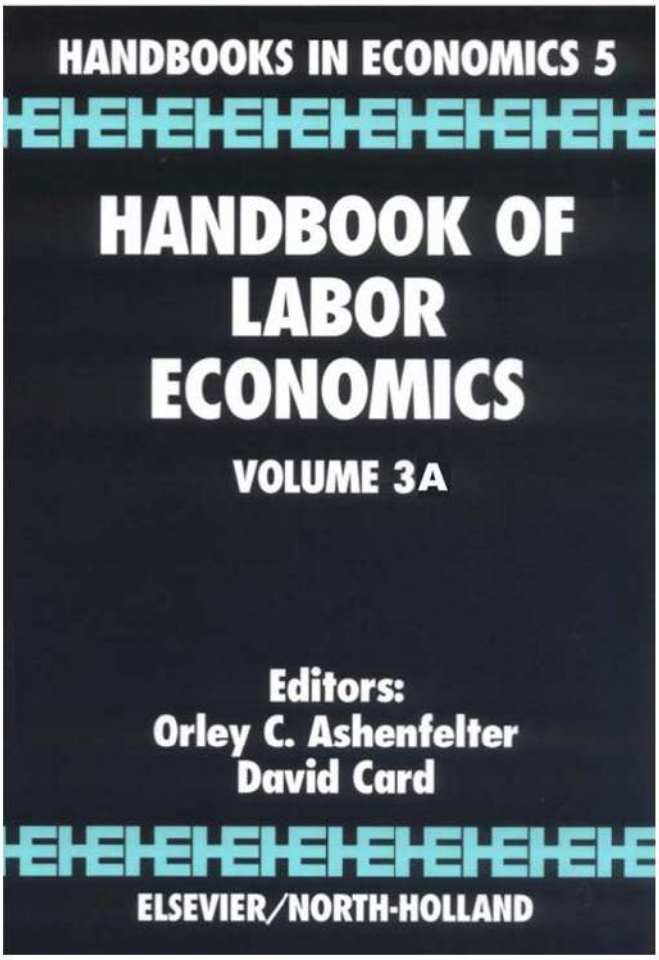

VOLUME 3A

PART 6 – OVERVIEW ISSUES

Chapter 23
Empirical Strategies in Labor Economics
JOSHUA D. ANGRIST and ALAN B. KRUEGER

Chapter 24
New Developments in Econometric Methods for Labor Market Analysis
ROBERT A. MOFFITT

Chapter 25
Institutions and Laws in the Labor Market
FRANCINE D. BLAU and LAWRENCE M. KAHN

Chapter 26
Changes in the Wage Structure and Earnings Inequality
LAWRENCE F. KATZ and DAVID H. AUTOR

PART 7 – THE SUPPLY SIDE

Chapter 27
Labor Supply: a Review of Alternative Approaches
RICHARD BLUNDELL and THOMAS MACURDY

Chapter 28
The Economic Analysis of Immigration
GEORGE J. BORJAS

Chapter 29
Intergenerational Mobility in the Labor Market
GARY SOLON

Chapter 30
The Causal Effect of Education on Earnings
DAVID CARD

Chapter 31
The Economics and Econometrics of Active Labor Market Programs
JAMES J. HECKMAN, ROBERT J. LALONDE and JEFFREY A. SMITH

Chapter 23

EMPIRICAL STRATEGIES IN LABOR ECONOMICS

JOSHUA D. ANGRIST*

MIT and NBER

ALAN B. KRUEGER*

Princeton University and NBER

Contents

Abstract 1278
JEL codes 1278
1 Introduction 1278
2 Identification strategies for causal relationships 1282
2.1 The range of causal questions 1282
2.2 Identification in regression models 1284
2.3 Consequences of heterogeneity and non-linearity 1309
2.4 Refutability 1326
3 Data collection strategies 1329
3.1 Secondary datasets 1332
3.2 Primary data collection and survey methods 1335
3.3 Administrative data and record linkage 1338
3.4 Combining samples 1339
4 Measurement issues 1339
4.1 Measurement error models 1340
4.2 The extent of measurement error in labor data 1344
4.3 Weighting and allocated values 1352
5 Summary 1354
Appendix A 1355
A.1 Derivation of Eq. (9) in the text 1355
A.2 Derivation of Eq. (34) in the text 1355
A.3 Schooling in the 1990 Census 1357
References 1357

* We thank Eric Bettinger, Lucia Breierova, Kristen Harknett, Aaron Siskind, Diane Whitmore, Eric Wang, and Steve Wu for research assistance. For helpful comments and discussions we thank Alberto Abadie, Daron Acemoglu, Jere Behrman, David Card, Angus Deaton, Jeff Kling, Guido Imbens, Chris Mazingo, Steve Pischke, and Cecilia Rouse. Of course, errors and omissions are solely the work of the authors.

Handbook of Labor Economics, Volume 3, Edited by O. Ashenfelter and D. Card

Abstract

This chapter provides an overview of the methodological and practical issues that arise when estimating causal relationships that are of interest to labor economists. The subject matter includes identification, data collection, and measurement problems. Four identification strategies are discussed, and five empirical examples – the effects of schooling, unions, immigration, military service, and class size – illustrate the methodological points. In discussing each example, we adopt an experimentalist perspective that emphasizes the distinction between variables that have causal effects, control variables, and outcome variables. The chapter also discusses secondary datasets, primary data collection strategies, and administrative data. The section on measurement issues focuses on recent empirical examples, presents a summary of empirical findings on the reliability of key labor market data, and briefly reviews the role of survey sampling weights and the allocation of missing values in empirical research.

JEL codes: J00; J31; C10; C81

1. Introduction

Empirical analysis is more common and relies on more diverse sources of data in labor economics than in economics more generally. Table 1, which updates Stafford's (1986, Table 7.2) survey of research in labor economics, bears out this claim. Indeed, almost 80% of recent articles published in labor economics contain some empirical work, and a striking two-thirds analyzed micro data. In the 1970s, micro data became more common in studies of the labor market than time-series data, and by the mid-1990s the use of micro data outnumbered time-series data by a factor of over ten to one. The use of micro and time-series data is more evenly split in other fields of economics.

In addition to using micro data more often, labor economists have come to rely on a wider range of datasets than other economists. The fraction of published papers using data other than what is in standard public-use files reached 38% in the period from 1994 to 1997. The files in the "all other micro datasets" category in Table 1 include primary datasets collected by individual researchers, customized public use files, administrative records, and administrative-survey links. This is noteworthy because about 10 years ago, in his *Handbook of Econometrics* survey of economic data issues, Griliches (1986, p. 1466) observed:

> ... since it is the 'badness' of the data that provides us with our living, perhaps it is not at all surprising that we have shown little interest in improving it, in getting involved in the grubby task of designing and collecting original datasets of our own.

The growing list of papers involving some sort of original data collection suggests this situation may be changing; examples include Freeman and Hall (1986), Ashenfelter and Krueger (1994), Anderson and Meyer (1994), Card and Krueger (1994, 1998), Dominitz and Manski (1997), Imbens et al. (1997), and Angrist (1998).

Labor economics has also come to be distinguished by the use of cutting edge econo-

Table 1
Percent of articles in each category[a]

	Labor economics articles					All fields
	1965–1969	1970–1974	1975–1979	1980–1983	1994–1997	1994–1997
Theory only	14	19	23	29	21	44
Micro data	11	27	45	46	66	28
Panel	1	6	21	18	31	12
Experiment	0	0	2	2	2	3
Cross-section	10	21	21	26	25	9
Micro dataset						
PSID	0	0	6	7	7	2
NLS	0	3	10	6	11	2
CPS	0	1	5	6	8	2
SEO	0	4	4	0	1	0
Census	3	5	2	0	5	1
All other micro datasets	8	14	18	27	38	21
Time series	42	27	18	16	6	19
Census tract	3	2	4	3	0	0
State	7	6	3	3	2	2
Other aggregate cross-section	14	16	8	4	6	6
Secondary data analysis	14	3	3	4	2	2
Total number of articles	106	191	257	205	197	993

[a] Notes: Figures for 1965–1983 are from Stafford (1986). Figures for 1994–1997 are based on the authors' analysis, and pertain to the first half of 1997. Following Stafford, articles are drawn from 8 leading economics journals.

metric and statistical methods. This claim is supported by the observation that outside of time-series econometrics, many and perhaps most innovations in econometric technique and style since the 1970s were motivated largely by research on labor-related topics. These innovations include sample selection models, non-parametric methods for censored data and survival analysis, quantile regression, and the renewed interest in statistical and identification problems related to instrumental variables estimators and quasi-experimental methods.

What do labor economists do with all the data they analyze? A broad distinction can be made between two types of empirical research in labor economics: descriptive analysis and causal inference. Descriptive analysis can establish facts about the labor market that need to be explained by theoretical reasoning and yield new insights into economic trends. The importance of ostensibly mundane descriptive analysis is captured by Sherlock Holmes's admonition that: "It is a capital offense to theorize before all the facts are in." A great deal of important research falls under the descriptive heading, including work on trends in poverty rates, labor force participation, and wage levels. A good

example of descriptive research of major importance is the work documenting the increase in wage dispersion in the 1980s (see e.g., Levy, 1987; Katz and Murphy, 1992; Murphy and Welch, 1992; Juhn et al., 1993). This research has inspired a vigorous search for the causes of changes in the wage distribution.

In contrast with descriptive analysis, causal inference seeks to determine the effects of particular interventions or policies, or to estimate features of the behavioral relationships suggested by economic theory. Causal inference and descriptive analysis are not competing methods; indeed, they are often complementary. In the example mentioned above, compelling evidence that wage dispersion increased in the 1980s inspired a search for causes of these changes. Causal inference is often more difficult than descriptive analysis, and consequently more controversial.

Most labor economists seem to share a common view of the importance of descriptive research, but there are differences in views regarding the role economic theory can or should play in causal modeling. This division is illustrated by the debate over social experimentation (Burtless, 1995; Heckman and Smith, 1995), in contrasting approaches to studying the impact of immigration on the earnings of natives (Card, 1990; Borjas et al., 1997), and in recent symposia illustrating alternative research styles (Angrist, 1995a; Keane and Wolpin, 1997). Research in a structuralist style relies heavily on economic theory to guide empirical work or to make predictions. Keane and Wolpin (1997, p. 111) describe the structural approach as trying to do one of two things: (a) recover the primitives of economic theory (parameters determining preferences and technology); (b) estimate decision rules derived from economic models. Given success in either of these endeavors, it is usually clear how to make causal statements and to generalize from the specific relationships and populations studied in any particular application.

An alternative to structural modeling, often called the quasi-experimental or simply the "experimentalist" approach, also uses economic theory to frame causal questions. But this approach puts front and center the problem of identifying the causal effects from specific events or situations. The problem of generalization of findings is often left to be tackled later, perhaps with the aid of economic theory or informal reasoning. Often this process involves the analysis of additional quasi-experiments, as in recent work on the returns to schooling (see, e.g., the papers surveyed by Card in this volume). In his methodological survey, Meyer (1995) describes quasi-experimental research as "an outburst of work in economics that adopts the language and conceptual framework of randomized experiments." Here, the ideal research design is explicitly taken to be a randomized trial and the observational study is offered as an attempt to approximate the force of evidence generated by an actual experiment.

In either a structural or quasi-experimental framework, the researcher's task is to estimate features of the causal relationships of interest. This chapter focuses on the *empirical strategies* commonly used to estimate features of the causal relationships that are of interest to labor economists. The chapter provides an overview of the methodological and practical issues that arise in implementing an empirical strategy. We use the term empirical strategy broadly, beginning with the statement of a causal question, and extend-

ing to identification strategies and econometric methods, selection of data sources, measurement issues, and sensitivity tests. The choice of topics was guided by our own experiences as empirical researchers and our research interests. As far as econometric methods go, however, our overview is especially selective; for the most part we ignore structural modeling since that topic is well covered elsewhere.[1] Of course, there is considerable overlap between structural and quasi-experimental approaches to causal modeling, especially when it comes to data and measurement issues. The difference is primarily one of emphasis, because structural modeling generally incorporates some assumptions about exogenous variability in certain variables and quasi-experimental analyses require some theoretical assumptions.

The attention we devote to quasi-experimental methods is also motivated by skepticism about the credibility of empirical research in economics. For example, in a critique of the practice of modern econometrics, Lester Thurow (1983, pp. 106–107) argued:

> Economic theory almost never specifies what secondary variables (other than the primary ones under investigation) should be held constant in order to isolate the primary effects. ... When we look at the impact of education on individual earnings, what else should be held constant: IQ, work effort, occupational choice, family background? Economic theory does not say. Yet the coefficients of the primary variables almost always depend on precisely what other variables are entered in the equation to "hold everything else constant."

This view of applied research strikes us as being overly pessimistic, but we agree with the focus on omitted variables. In labor economics, at least, the current popularity of quasi-experiments stems precisely from this concern: because it is typically impossible to adequately control for all relevant variables, it is often desirable to seek situations where it is reasonable to presume that the omitted variables are uncorrelated with the variables of interest. Such situations may arise if the researcher can use random assignment, or if the forces of nature or human institutions provide something close to random assignment.

The next section reviews four identification strategies that are commonly used to answer causal questions in contemporary labor economics. Five empirical examples – the effects of schooling, unions, immigration, military service, and class size – illustrate the methodological points throughout the chapter. In keeping with our experimentalist perspective, we attempt to draw clear distinctions between variables that have causal effects, control variables, and outcome variables in each example.

In Section 3 we turn to a discussion of secondary datasets and primary data collection strategies. The focus here is on data for the United States.[2] Section 3 also offers a brief review of issues that arise when conducting an original survey and suggestions for assem-

[1] See, for example, Heckman and MaCurdy's (1986) Handbook of Econometrics chapter, which "outlines the econometric framework developed by labor economists who have built theoretically motivated models to explain the new data." (p. 1918). We also have little to say about descriptive analysis because descriptive statistics are commonly discussed in statistics courses and books (see, e.g., Tukey, 1977; Tufte, 1992).

bling administrative datasets. Because existing public-use datasets have already been extensively analyzed, primary data collection is likely to be a growth industry for labor economists in the future. Following the discussion of datasets, Section 4 discusses measurement issues, including a brief review of classical models for measurement error and some extensions. Since most of this theoretical material is covered elsewhere, including the Griliches (1986) chapter mentioned previously, our focus is on topics of special interest to labor economists. This section also presents a summary of empirical findings on the reliability of labor market data, and reviews the role of survey sampling weights and the allocation of missing values in empirical research.

2. Identification strategies for causal relationships

> The object of science is the discovery of relations... of which the complex may be deduced from the simple. John Pringle Nichol, 1840 (quoted in Lord Kelvin's class notes).

2.1. The range of causal questions

The most challenging empirical questions in economics involve "what if" statements about counterfactual outcomes. Classic examples of "what if" questions in labor market research concern the effects of career decisions like college attendance, union membership, and military service. Interest in these questions is motivated by immediate policy concerns, theoretical considerations, and problems facing individual decision makers. For example, policy makers would like to know whether military cutbacks will reduce the earnings of minority men who have traditionally seen military service as a major career opportunity. Additionally, many new high school graduates would like to know what the consequences of serving in the military are likely to be for them. Finally, the theory of on-the-job training generates predictions about the relationship between time spent serving in the military and civilian earnings.

Regardless of the motivation for studying the effects of career decisions, the causal relationships at the heart of these questions involve comparisons of counterfactual states of the world. Someone – the government, an individual decision maker, or an academic economist – would like to know what outcomes would have been observed if a variable were manipulated or changed in some way. Lewis's (1986) study of the effects of union wage effects gives a concise description of this type of inference problem (p. 2): "At any given date and set of working conditions, there is for each worker a *pair* of wage figures, one for unionized status and the other for non-union status". Differences in these two

[2] Overviews of data sources for developing countries appear in Deaton's (1995) chapter in The Handbook of Development Economics, Grosh and Glewwe (1996, 1998), and Kremer (1997). We are not aware of a comprehensive survey of micro datasets for labor market research in Europe, though a few sources and studies are referenced in Westergard-Nielsen (1989).

potential outcomes define the causal effects of interest in Lewis's work, which uses regression to estimate the average gap between them.[3]

At first glance, the idea of unobserved potential outcomes seems straightforward, but in practice it is not always clear exactly how to define a counterfactual world. In the case of union status, for example, the counterfactual is likely to be ambiguous. Is the effect defined relative to a world where unionization rates are what they are now, a world where everyone is unionized, a world where everyone in the worker's firm or industry is unionized, or a world where no one is unionized? Simple micro-economic analysis suggests that the answers to these questions differ. This point is at the heart of Lewis's (1986) distinction between union *wage gaps*, which refers to causal effects on individuals, and *wage gains*, which refers to comparisons of equilibria in a world with and without unions. In practice, however, the problem of ambiguous counterfactuals is typically resolved by focusing on the consequences of hypothetical manipulations in the world as is, i.e., assuming there are no general equilibrium effects.[4]

Even if ambiguities in the definition of counterfactual states can be resolved, it is still difficult to learn about differences in counterfactual outcomes because the outcome of one scenario is all that is ever observed for any one unit of observation (e.g., a person, state, or firm). Given this basic difficulty, how do researchers learn about counterfactual states of the world in practice? In many fields, and especially in medical research, the prevailing view is that the best evidence about counterfactuals is generated by randomized trials because randomization ensures that outcomes in the control group really do capture the counterfactual for a treatment group. Thus, Federal guidelines for a new drug application require that efficacy and safety be assessed by randomly assigning the drug being studied or a placebo to treatment and control groups (Center for Drug Evaluation and Research, 1988). Leamer (1982) suggested that the absence of randomization is the main reason why econometric research often appears less convincing than research in other more experimental sciences. Randomized trials are certainly rarer in economics than in medical research, but labor economists are increasingly likely to use randomization to study the effects of labor market interventions (Passell, 1992). In fact, a recent survey of economists by Fuchs et al. (1998) finds that most labor economists place more credence in studies of the effect of government training programs on participants' income if the research design entails random assignment than if the research design is based on structural modeling.

Unfortunately, economists rarely have the opportunity to randomize variables like educational attainment, immigration, or minimum wages. Empirical researchers must therefore rely on observational studies that typically fail to generate the same force of evidence as a randomized experiment. But the object of an observational study, like an experimental study, can still be to make comparisons that provide evidence about causal

[3] See also Rubin (1974, 1977) and Holland (1986) for formal discussions of counterfactual outcomes in causal research.

[4] Lewis's (1963) earlier book discussed causal effects in terms of industries and sectors, and made a distinction between "direct" and "indirect" effects of unions similar to the distinction between wage gaps and wage gains. Heckman et al. (1998) discuss general equilibrium effects that arise in the evaluation of college tuition subsidies.

effects. Observational studies attempt to accomplish this by controlling for observable differences between comparison groups using regression or matching techniques, using pre-post comparisons on the same units of observation to reduce bias from unobserved differences, and by using instrumental variables as a source of quasi-experimental variation. Randomized trials form a conceptual benchmark for assessing the success or failure of observational study designs that make use of these ideas, even when it is clear that it may be impossible or at least impractical to study some questions using random assignment. In almost every observational study, it makes sense to ask whether the research design is a good "natural experiment."[5]

A sampling of causal questions that economists have studied without benefit of a randomized experiment appears in Table 2, which characterizes a few observational studies grouped according to the source of variation used to make causal inferences about a single "causing variable." The distinction between causing variables and control variables in Table 2 is one difference between the discussion in this chapter and traditional econometric texts, which tend to treat all variables symmetrically. The combination of a clearly labeled source of identifying variation in a causal variable and the use of a particular econometric technique to exploit this information is what we call an *identification strategy*. Studies were selected for Table 2 primarily because the source or type of variation that is being used to make causal statements is clearly labeled. The four approaches to identification described in the table are: Control for Confounding Variables, Fixed-effects and Differences-in-differences, Instrumental Variables, and Regression Discontinuity methods. This taxonomy provides an outline for the next section.

2.2. *Identification in regression models*

2.2.1. *Control for confounding variables*

Labor economists have long been concerned with the question of whether the positive association between schooling and earnings is a causal relationship. This question originates partly in the observation that people with more schooling appear to have other characteristics, such as wealthier parents, that are also associated with higher earnings. Also, the theory of human capital identifies unobserved earnings potential or "ability" as one of the principal determinants of educational attainment (see, e.g, Willis and Rosen, 1979). The most common identification strategy in research on schooling (and in economics in general) attempts to reduce bias in naive comparisons by using regression to control

[5] This point is also made by Freeman (1989). The notion that experimentation is an ideal research design for Economics goes back at least to the Cowles Commission. See, for example, Girshick and Haavelmo (1947), who wrote (p. 79): "In economic theory ... the total demand for the commodity may be considered a function of all prices and of total disposable income of all consumers. The ideal method of verifying this hypothesis and obtaining a picture of the demand function involved would be to conduct a large-scale experiment, imposing alternative prices and levels of income on the consumers and studying their reactions." Griliches and Mairesse (1998, p. 404) recently argued that the search for better natural experiments should be a cornerstone of research on production functions.

for variables that are confounded with (i.e., related to) schooling. The typical estimating equation in this context is,

$$Y_i = X_i'\beta_r + \rho_r S_i + e_i, \tag{1}$$

where Y_i is person i's log wage or earnings, X_i is a $k \times 1$ vector of control variables, including measures of ability and family background, S_i is years of educational attainment, and e_i is the regression error. The vector of population parameters is $[\beta_r' \ \rho_r]'$. The "r" subscript on the parameters signifies that these are *regression* coefficients. The question of causality concerns the interpretation of these coefficients. For example, they can always be viewed as providing the best (i.e., minimum-mean-squared-error) linear predictor of Y_i.[6] The best linear predictor need not have causal or behavioral significance; the resulting residual is uncorrelated with the regressors simply because the first-order conditions for the prediction problem are $E[e_i X_i] = 0$ and $E[e_i S_i] = 0$.

Regression estimates from five early studies of the relationship between schooling, ability, and earnings are summarized in Table 3. The first row reports estimates without ability controls while the second row reports estimates that include some kind of test score in the X-vector as a control for ability. Information about the X-variables is given in the rows labeled "ability variable" and "other controls". The first two studies, Ashenfelter and Mooney (1968) and Hansen et al. (1970) use data on individuals at the extremes of the ability distribution (graduate students and military rejects), while the others use more representative samples. Results from the last two studies, Griliches and Mason (1972) and Chamberlain (1978), are reported for models with and without family background controls.

The schooling coefficients in Table 3 are smaller than the coefficient estimates we are used to seeing in studies using more recent data (see, e.g., Card's survey in this volume). This is partly because the association between earnings and schooling has increased, partly because the samples used in the papers summarized in the table include only young men, and partly because the models used for estimation control for age and not potential experience (age-education-6). The latter parameterization leads to larger coefficient estimates since, in a linear model, the schooling coefficient controlling for age is equal to the schooling coefficient controlling for experience minus the experience coefficient. The only specification in Table 2 that controls for potential experience is from Griliches (1977), which also generates the highest estimate in the table (0.065). The corresponding estimate controlling for age is 0.022. The table also shows that controlling for ability and family background generally reduces the magnitude of schooling coefficients, implying that at least some of the association between earnings and schooling in these studies can be attributed to variables other than schooling.

What conditions must be met for regression estimates like those in Table 3 to have a

[6] The best linear predictor is the solution to $\text{Min}_{b,c} E[(Y_i - X_i'b - cS_i)^2]$ (see, e.g., White, 1980; Goldberger, 1991).

Table 2
Identification strategies in observational studies[a]

Type of identifying information	Outcome variable	Causing variable	Estimator	Reference
I. Control for confounding variables				
Control for ability and family background	Wages	Years of schooling	Regression	See Table 3
Control for past outcomes	Employment	Training programs	Regression and matching	Card and Sullivan (1988)
	Earnings		Propensity score matching	Dehejia and Wahba (1995)
			Propensity score matching	Heckman et al. (1997)
Control for military selection criteria	Earnings	Veteran status	Regression and matching	Angrist (1998)
II. Fixed-effects and differences-in-differences				
Panel data/individual changes in status	Wages	Union status	Differencing/ analysis of covariance	Freeman (1984)
	Earnings	Training programs	Differences-in-differences	Ashenfelter and Card (1985)
The Mariel Boatlift	Employment of natives	Numbers of immigrants	Differences-in-differences	Card (1990)
Changes in state laws or rules	Injury duration	Workers' Compensation benefit	Differences-in-differences	Meyer et al. (1995)
	Unemployment	Unemployment	Differences-in-differences	Solon (1985)
	Duration	Insurance benefit	Hazard models	

Change in Federal law	Employment	Anti-discrimination policy	Differences-in-differences	Heckman and Payner (1989)
Twin comparisons	Income	Years of schooling	Differencing	Behrman et al. (1980); Taubman (1976)
	Earnings		Differencing/IV	Ashenfelter and Krueger (1994)
III. Instrumental variables				
Twin births	Schooling	Fertility Teen fertility	2SLS	Rozensweig and Wolpin (1980), Bronars and Grogger (1994)
Twin births Sibling-sex composition	Labor supply	Fertility		Angrist and Evans (1998)
Year of birth Quarter of birth	Wages	Years of schooling	2SLS	Hausman and Taylor (1981) Angrist and Krueger (1991)
Draft lottery Year of birth	Earnings	Veteran status	Two-sample IV	Angrist (1990) Imbens and van der Klaauw (1995)
IV. Regression-discontinuity methods				
Financial aid thresholds	College enrollment	Financial aid	2SLS	van der Klaauw (1996)
Class-size maximum	Test scores	Class size	2SLS	Angrist and Lavy (1998)
Social Security Notch	Labor force participation	Social Security benefits	OLS	Krueger and Pischke (1992)

[a] Notes: The table lists studies classified by type of identification strategy.

Table 3
Effects of ability controls on estimates of the returns to education – early studies

	AM-68 (1)	HWC-70 (2)	G-77 (3)	G-77 (4)	GM-72 (5)	GM-72 (6)	C-78 (7)	C-78 (8)
A. OLS estimates[a]								
No ability controls	0.0205	0.0346	0.022	0.065	0.0508	0.0408	0.041	0.036
	(0.0285)	(0.007)	(0.005)	(0.005)	(0.0039)	(0.0041)	(0.006)	(0.007)
Ability controls	0.0213	0.0171	0.014	0.059	0.0433	0.0365	0.030	0.027
		(0.0074)	(0.005)	(0.005)	(0.0044)	(0.0046)	(0.007)	(0.008)
Ability variable	Math SAT	AFQT	IQ	IQ	AFQT	AFQT	IQ	IQ
Other controls	Field, degree, occupation, experience	Age/race	Age	Experience	Age, length of military service	Age, length of military service, region/urban controls, father's occupation + schooling	Age, race, region/urban controls	Age, race, region/urban controls plus family background variables
Dependant variable	Annual earnings in 1966	Annual income in 1962 (minus transfers)	Hourly wage in 1969	Hourly wage in 1969	Usual weekly earnings in 1964	Usual weekly earnings in 1964	Hourly wage in 1973	Hourly wage in 1973
Data set	Male Ph.D.'s beginning in 1958–1960	Rejected low-AFQT military applicants	NLSYM	NLSYM	1964 CPS veterans	1964 CPS veterans	NLSYM	NLSYM
N	694	2403	1362	1362	1454	1454	897	897
B. Adjusting for endogenous ability and measurement error in schooling[b]								
				0.085				0.132
				(0.009)				(0.049)

[a] Except for column (1), these are OLS schooling coefficients in semilog wage equations. Studies are Ashenfelter and Mooney (1968), Hansen et al. (1970), Griliches (1977), Griliches and Mason (1972), and Chamberlain (1978). Coefficients and standard errors in column (1) were constructed from coefficients on single year-of-education dummies, averaged and divided by average earnings. Studies were selected for the table because they report comparable specifications with and without ability controls.

[b] Estimates include the same controls as in (A). The Griliches estimates use instrumental variables and treat schooling and IQ as endogenous. The Chamberlain estimates are from a structural model that allows for measurement error and unobserved ability. The Chamberlain dependent variable in the structural model is the hourly wage in 1971.

causal interpretation? In this case, causality can be based on an underlying functional relationship that describes what a given individual would earn if he or she obtained different levels of education. This relationship may be person-specific, so we write

$$Y_{S,i} \equiv f_i(S) \tag{2}$$

to denote the potential (or latent) earnings that person i would receive after obtaining S years of education. Note that the function $f_i(S)$ has an i subscript on it while S does not. This highlights the fact that although S is a variable, it is not a random variable. The function $f_i(S)$ tells us what *i would earn* for any value of schooling, S, and not just for the realized value, S_i. In other words, $f_i(S)$ answers "what if" questions. In the context of theoretical models of the relationship between human capital and earnings, the form of $f_i(S)$ may be determined by aspects of individual behavior and/or market forces. With or without an explicit economic model for $f_i(S)$, however, we can think of this function as describing the earnings level of individual i if that person were assigned schooling level S (e.g., in an experiment).

Once the causal relationship of interest, $f_i(S)$, has been defined, it can be linked to the observed association between schooling and earnings. A convenient way to do this is with a linear model:

$$f_i(S) = \beta_0 + \rho S + \eta_i. \tag{3}$$

In addition to being linear, this equation says that the functional relationship of interest is the same for all individuals. Again, S is written without a subscript, because Eq. (3) tells us what person i would earn for any value of S and not just the realized value, S_i. The only individual-specific and random part of $f_i(S)$ is a mean-zero error component, η_i, which captures unobserved factors that determine earnings. In practice, regression estimates have a causal interpretation under weaker functional-form assumptions than this but we postpone a detailed discussion of this point until Section 2.3. Note that the earnings of someone with no schooling at all is just $\beta_0 + \eta_i$ in this model.

Substituting the observed value S_i for S in Eq. (3), we have

$$Y_i = \beta_0 + \rho S_i + \eta_i. \tag{4}$$

This looks like Eq. (1) without covariates, except that Eq. (3) explicitly associates the regression coefficients in Eq. (4) with a causal relationship. The OLS estimate of ρ in Eq. (4) has probability limit

$$C(Y_i, S_i)/V(S_i) = \rho + C(S_i, \eta_i)/V(S_i). \tag{5}$$

The term $C(S_i, \eta_i)/V(S_i)$ is the coefficient from a regression of η_i on S_i, and reflects any correlation between the realized S_i and unobserved individual earnings potential, which in this case is the same as correlation with η_i. If educational attainment were randomly assigned, as in an experiment, then we would have $C(S_i, \eta_i) = 0$ in the linear model. In practice, however, schooling is a consequence of individual decisions and institutional

forces that are likely to generate correlation between η_i and schooling. Consequently, it is not automatic that OLS provides a consistent estimate of the parameter of interest.[7]

Regression strategies attempt to overcome this problem in a very simple way: in addition to the functional form assumption for potential outcomes embodied in (3), the random part of individual earnings potential, η_i, is decomposed into a linear function of the k observable characteristics, X_i, and an error term, ε_i,

$$\eta_i = X_i'\beta + \varepsilon_i, \tag{6a}$$

where β is a vector of population regression coefficients. This means that ε_i and X_i are uncorrelated by construction. The key identifying assumption is that the observable characteristics, X_i, are the *only* reason why η_i and S_i (equivalently, $f_i(S)$ and S_i) are correlated, so

$$\mathrm{E}[S_i\varepsilon_i] = 0. \tag{6b}$$

This is the "selection on observables" assumption discussed by Barnow et al. (1981), where the regressor of interest is assumed to be determined independently of potential outcomes after accounting for a set of observable characteristics.

Continuing to maintain the selection-on-observables assumption, a consequence of (6a) and (6b) is that

$$C(Y_i, S_i)/V(S_i) = \rho + \Gamma_{SX}'\beta, \tag{7}$$

where Γ_{SX} is a $k \times 1$ vector coefficients from a regression of each element of X_i on S_i. Eq. (7) is the well known "omitted variables bias" formula, which relates a bivariate regression coefficient to the coefficient on S_i in a regression that includes additional covariates. If the omitted variables are positively related to earnings ($\beta > 0$) and positively correlated with schooling ($\Gamma_{SX} > 0$), then $C(Y_i, S_i)/V(S_i)$ is larger than the causal effect of schooling, ρ. A second consequence of (6a) and (6b) is that the OLS estimate of ρ_r in Eq. (1) is in fact consistent for the causal parameter, ρ. Note, however, that in this discussion of the problem of causal inference, $\mathrm{E}[S_i\varepsilon_i] = 0$ is an *assumption* about ε_i and S_i, whereas $\mathrm{E}[X_i\varepsilon_i] = 0$ is a statement about covariates that is true by *definition*. This suggests that it is important to distinguish error terms that represent the random parts of models for potential outcomes from mechanical decompositions where the relationship between errors and regressors has no behavioral content.

A key question in any regression study is whether the selection-on-observables assumption is plausible. This assumption clearly makes sense when there is actual random assignment conditional on X_i. Even without random assignment, however, selection-on-observables might be plausible if we know a lot about the process generating the regressor of interest. We might know, for example, that applicants to a particular college or univer-

[7] Econometric textbooks (e.g., Pindyk and Rubinfeld, 1991) sometimes refer to regression models for causal relationships as "true models," but this seems like potentially misleading terminology since non-behavioral descriptive regressions could also be described as being "true".

sity are screened using certain characteristics, but conditional on these characteristics all applicants are acceptable and chosen on a first-come/first-serve basis. This leads to a situation like the one described by Barnow et al. (1981, p. 47), where "Unbiasedness is attainable when the variables that determined the assignment are known, quantified, and included in the equation." Similarly, Angrist (1998) argued that because the military is known to screen applicants on the basis of observed characteristics, comparisons of veteran and non-veteran applicants that adjust for these characteristics have a causal interpretation. The case for selection-on-observables in a generic schooling equation is less clear cut, which is why so much attention has focused on the question of omitted-variables bias in OLS estimates of schooling coefficients.

Regression pitfalls. Schooling is not randomly assigned and, as in many other problems, we do not have detailed institutional knowledge about the process that actually determines assignment. The choice of covariates is therefore crucial. Obvious candidates include any variables that are correlated with both schooling and earnings. Test scores are good candidates because many educational institutions use tests to determine admissions and financial aid. On the other hand, it is doubtful that any particular test score is a perfect control for all the differences in earnings potential between more and less educated individuals. We see this in the fact that adding family background variables like parental income further reduces the size of schooling coefficients. A natural question about any regression control strategy is whether the estimates are highly sensitive to the inclusion of additional control variables. While one should always be wary of drawing causal inferences from observational data, sensitivity of regression results to changes in the set of control variables is an extra reason to wonder whether there might be unobserved covariates that would change the estimates even further.

The previous discussion suggests that Table 3 can be interpreted as showing that there is significant ability bias in OLS estimates of the causal effect of schooling on earnings. On the other hand, a number of concerns less obvious than omitted-variables bias suggest this conclusion may be premature. A theme of the Griliches and Chamberlain papers cited in the table is that the negative impact of ability measures on schooling coefficients is eliminated and even reversed after accounting for two factors: measurement error in the regressor of interest, and the use of endogenous test score controls that are themselves *affected by* schooling.

A standard result in the analysis of measurement error is that if variables are measured with an additive error that is uncorrelated with correctly-measured values, this imparts an attenuation bias that shrinks OLS estimates towards zero (see, e.g., Griliches, 1986; Fuller, 1987, and Section 4). The proportionate reduction is one minus the ratio of the variance of correctly-measured values to the variance of measured values. Furthermore, the inclusion of control variables that are correlated with actual values and uncorrelated with the measurement error tends to aggravate this attenuation bias. The intuition for this result is that the residual variance of true values is reduced by the inclusion of additional control variables while the residual variance of the measurement error is left unchanged. Although

studies of measurement error in education data suggest that only 10% of the variance in measured education is attributable to measurement error, it turns out that the downward bias in regression models with ability and other controls can still be substantial.[8]

A second complication raised in the early literature on regression estimates of the returns to schooling is that variables used to control for ability may be endogenous (see, e.g., Griliches and Mason, 1972, or Chamberlain, 1977). If wages and test scores are *both* outcomes that are affected by schooling, then test scores cannot play the role of an exogenous, pre-determined control variable in a wage equation. To see this, consider a simple example where the causal relationship of interest is (4), and $C(S_i, \eta_i) = 0$ so that a bivariate regression would in fact generate a consistent estimate of the causal effect. Suppose that schooling affects test scores as well as earnings, and that the effect on test scores can be expressed using the model

$$A_i = \gamma_0 + \gamma_1 S_i + \eta_{1i}. \tag{8}$$

This relationship can be interpreted as reflecting the fact that more formal schooling tends to improve test scores (so $\gamma_1 > 0$). We also assume that $C(S_i, \eta_{1i}) = 0$, so that OLS estimates of (8) would be consistent for γ_1. The question is what happens if we add the outcome variable, A_i, to the schooling equation in a mistaken (in this case) attempt to control for ability bias.

Endogeneity of A_i in this context means that η_i and η_{1i} are correlated. Since people who do well on standardized tests probably earn more for reasons other than the fact that they have more schooling, it seems reasonable to assume that $C(\eta_i, \eta_{1i}) > 0$. In this case, the coefficient on S_i in a regression of Y_i on S_i and A_i leads to an inconsistent estimate of the effect of schooling. Evaluation of probability limits shows that the OLS estimate of the schooling coefficient in a model that includes A_i converges to

$$C(Y_i, S_{Ai})/V(S_{Ai}) = \rho - \gamma_1 \varphi_{01}, \tag{9}$$

where S_{Ai} is the residual from a regression of S_i on A_i and φ_{01} is the coefficient from a regression of η_i on η_{1i} (see Appendix A for details). Since $\gamma_1 > 0$ and $\varphi_{01} > 0$, controlling for the endogenous test score variable tends to make the estimate of the returns to schooling smaller, but this is not because of any omitted-variables bias in the equation of interest. Rather it is a consequence of the bias induced by conditioning on an outcome variable.[9]

The problems of measurement error and endogenous regressors generate identification challenges that lead researchers to use methods beyond the simple regression-control framework. The most commonly employed strategies for dealing with these problems

[8] For a detailed elaboration of this point, see Welch (1975) or Griliches (1977), who notes (p. 13): "Clearly, the more variables we put into the equation which are related to the systematic components of schooling, and the better we 'protect' ourselves against various possible biases, the worse we make the errors of measurement problem." We present some new evidence on attenuation and covariates in Section 4.

[9] A similar problem may affect estimates of schooling coefficients in equations that control for occupation. Like test scores and other ability measures, occupation is itself a consequence of schooling that is probably correlated with unobserved earnings potential. For a related discussion of matching estimates, see Rosenbaum (1984).

involve instrumental variables (IV), two-stage least squares (2SLS), and latent-variable models. We briefly mention some 2SLS and latent-variable estimates, but defer a detailed discussion of 2SLS and related IV strategies until Section 2.2.3. The major practical problem in models of this type is to find valid instruments for schooling and ability. Panel B reports Griliches (1977) 2SLS estimates of Eq. (1) treating both schooling and IQ scores as endogenous. The instruments are family background measures and a second ability proxy. Chamberlain (1978) develops an alternate approach that uses panel data to identify the effects of endogenous schooling in a latent-variable model for unobserved ability. Both the Chamberlain (1978) and Griliches (1977) estimates are considerably larger than the corresponding OLS estimates, a finding which led these authors to conclude that the empirical case for a negative ability bias in schooling coefficients is much weaker than the OLS estimates suggest.[10]

2.2.2. *Fixed effects and differences-in-differences*

The main idea behind fixed-effects identification strategies is to use repeated observations on individuals (or families) to control for *unobserved* and unchanging characteristics that are related to both outcomes and causing variables. A classic field of application for fixed-effects models is the attempt to estimate the effect of union status. Suppose, for example, that we would like to know the effect of workers' union status on their wages. That is, for each worker, we imagine that there are two potential outcomes, Y_{0i}, denoting what the worker would earn if not a union member, and Y_{1i} denoting what the worker would earn as a union member. This is just like $Y_{S,i}$ in the schooling example, except that here S is the dichotomous variable, union status. The effect of union status on an individual worker is $Y_{1i} - Y_{0i}$, but this is never observed directly since only one potential outcome is ever observed for each individual at any one time.[11]

Most analyses of the union problem begin with a constant-coefficients regression model for potential outcomes, where

$$Y_{0i} = X_i'\beta + \varepsilon_i, \qquad Y_{1i} = Y_{0i} + \delta. \tag{10}$$

As in the schooling problem, Y_{0i} has been decomposed into a linear function of observed covariates, $X_i'\beta$, and a residual, ε_i, that is uncorrelated with X_i by construction. Using U_i to indicate union members, this leads to the regression equation,

$$Y_i = X_i'\beta + U_i\delta + \varepsilon_i, \tag{11}$$

which describes the causal relationship of interest.

Many researchers working in this framework have argued that union status is likely to be related to potential non-union wages, Y_{0i}, even after conditioning on covariates, X_i (see,

[10] Another strand of the literature on causal effects of schooling uses sibling data to control for family effects that are shared by siblings; early studies are by Gorseline (1932) and Taubman (1976); see also Griliches' (1979) survey. Here the problem of measurement error is paramount (see Sections 2.2.2 and 4.1).

[11] This notation for counterfactual outcomes was used by Rubin (1974, 1977). Siegfried and Sweeney (1980) and Chamberlain (1980) use a similar notation to discuss the effect of a classroom intervention on test scores.

e.g., Abowd and Farber, 1982; or Chapters 4 and 5 in Lewis, 1986). This means that U_i is correlated with ε_i, so OLS does not estimate the causal effect, δ. An alternative to OLS uses panel datasets such as matched CPS rotation groups, the Panel Study of Income Dynamics, or the National Longitudinal Surveys, and exploits repeated observations on individuals to control for unobserved individual characteristics that are time-invariant. A well-known study in this genre is Freeman (1984).

The following model, similar to many in the literature on union status, illustrates the fixed-effects approach. Modifying the previous notation to incorporate $t = 1, \ldots, T$ observations on individuals, the fixed-effects solution for this problem begins by writing

$$Y_{0it} = X'_{it}\beta_t + \lambda\alpha_i + \xi_{it}, \tag{12}$$

where α_i is an unobserved variable for person i, that we could, in principle, include as a control if it were observed. Eq. (12) is a regression decomposition with covariates X_{it} and α_i, so ξ_{it} is uncorrelated with X_{it} and α_i by construction (X_{it} can include characteristics from different periods). The causal/regression model for panel data is now

$$Y_{it} = X'_{it}\beta_t + U_{it}\delta_t + \lambda\alpha_i + \xi_{it}, \tag{13}$$

where we have allowed the causal effect of interest to be time-varying. The identifying assumptions are that the coefficient λ does not vary across periods and that

$$E[U_{it}\xi_{is}] = 0 \qquad \text{for } s = 1, \ldots, T. \tag{14}$$

In other words, whatever the source of correlation is between U_{it} and unobserved earnings potential, it can be described by an additive time-invariant covariate α_i, that has the same coefficient each period. Since differencing eliminates $\lambda\alpha_i$, OLS estimates of the differenced equation

$$Y_{it} - Y_{it-k} = X'_{it}\beta_t - X'_{it-k}\beta_{t-k} + U_{it}\delta_t - U_{it-k}\delta_{t-k} + (\xi_{it} - \xi_{it-k}) \tag{15}$$

are consistent for the parameters of interest.

Any transformation of the data that eliminates the unobserved α_i can be used to estimate the parameters of interest in this model. One of the most popular estimators in this case is the deviations-from-means or the analysis of covariance (ANCOVA) estimator, which is most often used for models where β_t and δ_t are assumed to be fixed. The analysis of covariance estimator is OLS applied to

$$Y_{it} - \bar{y}_i = \beta'(X_{it} - \bar{x}_i) + \delta(U_{it} - \bar{u}_i) + (\xi_{it} - \bar{\xi}_i), \tag{16}$$

where overbars denote person-averages. Analysis of covariance is preferable to differencing on efficiency grounds in some cases; for models with normally distributed homoscedastic errors, ANCOVA is the maximum likelihood estimator. An alternative econometric strategy for the estimation of models with individual effects uses repeated observations on cohort averages instead of repeated data on individuals. For details and examples see Ashenfelter (1984) or Deaton (1985).

Finally, note that while standard fixed-effects estimators can only be used to estimate

the effects of time-varying regressors, Hausman and Taylor (1981) have developed a hybrid panel/IV procedure for models with time-invariant regressors (like schooling). It is also worth noting that even if the causing variable of interest is time-invariant, we can use standard fixed-effects estimators to estimate *changes* in the effect of a time invariant variable. For example, the estimating equation for a model with fixed U_i is

$$Y_{it} - Y_{it-k} = X'_{it}\beta_t - X'_{it-k}\beta_{t-k} + U_i(\delta_t - \delta_{t-k}) + (\xi_{it} - \xi_{it-k}), \tag{17}$$

so $(\delta_t - \delta_{t-k})$ is identified. Angrist (1995b) used this method to estimate changes in schooling coefficients in the West Bank and Gaza Strip even though schooling is approximately time-invariant.

Fixed-effects pitfalls. The use of panel data to eliminate bias from unobserved individual effects raises a number of econometric and statistical issues. Since this material is covered in Chamberlain's (1984) chapter in *The Handbook of Econometrics,* we limit our discussion to an overview of problems that have been of particular concern to labor economists. First, analysis of covariance and differencing estimators are not consistent when the process determining U_{it} involves lagged dependent variables. This issue comes up in the analysis of training programs because participants often experience a pre-program decline in earnings, a fact first noted by Ashenfelter (1978). If past earnings are observed and there are no unobserved individual effects, the simplest strategy is to control for past earnings either by including lagged earnings as a regressor or in matched treatment-control comparisons (see, e.g., Dehejia and Wahba, 1995; Heckman et al., 1997). In fact, the question of whether trainees and a candidate comparison group have similar lagged outcomes is sometimes seen as a litmus test for the legitimacy of the comparison group in the evaluation of training programs (see, e.g., Heckman and Hotz, 1989).

A problem arises in this context, however, when the process determining U_{it} involves past outcomes and an unobserved covariate, α_i. Ashenfelter and Card (1985) discuss an example involving the effect of training on the Social Security-taxable earnings of trainees under the Comprehensive Employment and Training Act (CETA). They propose a model of training status where individuals who enter CETA training in year τ do so because they have low α_i *and* their earnings were unusually low in year $\tau - 1$. Suppose initially we ignore the fact that training status involves past earnings, and estimate an equation like (15). Ignoring other covariates, this amounts to comparing the earnings growth of trainees and controls. But whatever the true program effect is, the growth in the earnings of CETA trainees from year $\tau - 1$ to year $\tau + 1$ will tend to be larger than the earnings growth in a candidate control group simply because of regression-to-the-mean. This generates a spurious positive training effect and the conventional differencing method breaks down.[12]

A natural strategy for dealing with this problem might seem to be to add $Y_{i\tau-1}$ to the list of control variables, and then difference away the fixed effect in a model with $Y_{i\tau-1}$ as regressor. The problem is that now any transformation that eliminates the fixed effect will

[12] Deviations-from-means estimators are also biased in this case.

leave at least one regressor – the lagged dependent variable – correlated with the errors in the transformed equation. Although the lagged dependent variable is not the regressor of interest, the fact that it is correlated with the error term in the transformed equation means that the estimate of the coefficient on $U_{i\tau+1}$ is biased as well. A detailed description of this problem, and the solutions that have been proposed for it, raises technical issues beyond the scope of this chapter. A useful reference is Nickell, 1981, especially pp. 1423–1424. See also Card and Sullivan's (1988) study of the effect of CETA training on the employment rates of trainees, which reports both fixed-effects estimates and matching estimates that control for lagged outcomes.

A second potential problem with fixed-effects estimators is that bias from measurement error is usually aggravated by transformations that eliminate the individual effects (see, e.g., Freeman, 1984; Griliches and Hausman, 1986). This fact may explain why fixed-effects estimates often turn out to be smaller than estimates in levels. Finally, perhaps the most important problem with this approach is that the assumption that omitted variables can be captured by an additive, time-invariant individual effect is arbitrary in the sense that it usually does not come from economic theory or from information about the relevant institutions.[13] On the other hand, the fixed-effects approach has intuitive appeal ("whatever makes us special is timeless") and an identification payoff that is hard to beat. Also, fixed-effects models lend themselves to a variety of specification tests. See, for example, Ashenfelter and Card (1985), Chamberlain (1984), Griliches and Hausman (1986), Angrist and Newey (1991), and Jakubson (1991). Many of these studies also focus on the union example.

The differences-in-differences (DD) model. Differences-in-differences strategies are simple panel-data methods applied to sets of group means in cases when certain groups are exposed to the causing variable of interest and others are not. This approach, which is transparent and often at least superficially plausible, is well-suited to estimating the effect of sharp changes in the economic environment or changes in government policy. The DD method has been used in hundreds of studies in economics, especially in the last two decades, but the basic idea has a long history. An early example in labor economics is Lester (1946), who used the differences-in-differences technique to study employment effects of minimum wages.[14]

The DD approach is explained here using Card's (1990) study of the effect of immigration on the employment of natives as an example. Some observers have argued that immigration is undesirable because low-skilled immigrants may displace low-skilled or less-educated US citizens in the labor market. Anecdotal evidence for this claim includes newspaper accounts of hostility between immigrants and natives in some cities, but the empirical evidence is inconclusive. See Friedberg and Hunt (1995) for a survey of research on this question. As in our earlier examples, the object of research on immigration is to

[13] An exception is the literature on life-cycle labor supply (e.g., MaCurdy, 1981; Altonji, 1986).

[14] The DD method goes by different names in different fields. Psychologist Campbell (1969) calls it the "nonequivalent control-group pretest-posttest design."

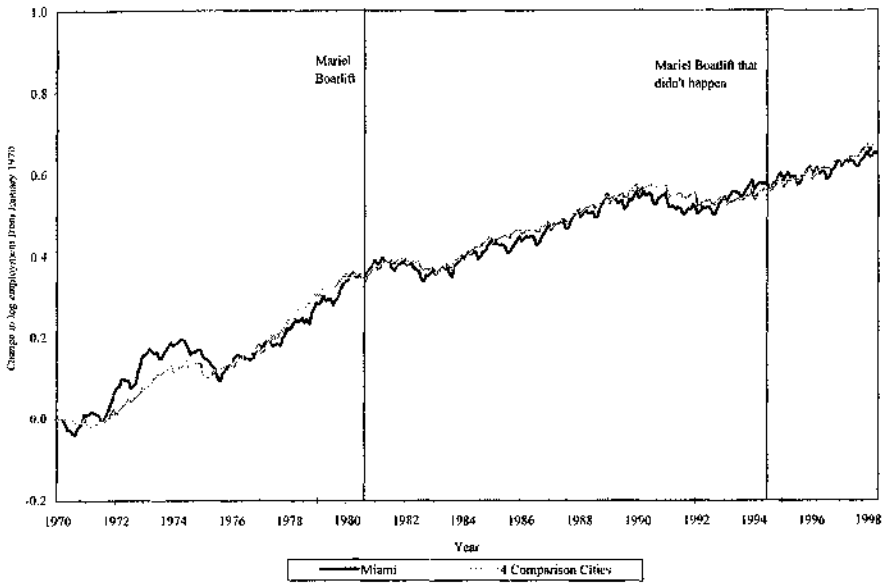

Fig. 1. Changes in employment in Miami and comparison cities. *Source*: authors' calculations from BLS State and Area Employment, Hours, and Earnings Establishment Survey.

find some sort of comparison that provides a compelling answer to "what if" questions about the consequences of immigration.

Card's study used a sudden large-scale migration from Cuba to Miami known as the Mariel Boatlift to make comparisons and answer counterfactual questions about the consequences of immigration. In particular, Card asks whether the Mariel immigration, which increased the Miami labor force by about 7% between May and September of 1980, reduced the employment or wages of non-immigrant groups. An important component of this identification strategy is the selection of comparison cities that can be used to estimate *what would have happened* in the Miami labor market absent the Mariel immigration.

The comparison cities Card used in the Mariel Boatlift study were Atlanta, Los Angeles, Houston, and Tampa-St. Petersburg. These cities were chosen because, like Miami, they have large Black and Hispanic populations and because discussions of the impact of immigrants often focuses on the consequences for minorities. Most importantly, these cities appear to have employment trends similar to those in Miami at least since 1976. This is documented in Fig. 1, which is similar to a figure in Card's (1989) working paper that did not appear in the published version of his study. The figure plots monthly observations on the log of employment in Miami and the four comparison cities from 1970 through 1998. The two series, which are from BLS establishment data, have been normalized by subtracting the 1970 value.

Table 4
Differences-in-differences estimates of the effect of immigration on unemployment[a]

	Group	Year		
		1979 (1)	1981 (2)	1981–1979 (3)
	Whites			
(1)	Miami	5.1 (1.1)	3.9 (0.9)	−1.2 (1.4)
(2)	Comparison cities	4.4 (0.3)	4.3 (0.3)	−0.1 (0.4)
(3)	Miami-Comparison Difference	0.7 (1.1)	−0.4 (0.95)	−1.1 (1.5)
	Blacks			
(4)	Miami	8.3 (1.7)	9.6 (1.8)	1.3 (2.5)
(5)	Comparison cities	10.3 (0.8)	12.6 (0.9)	2.3 (1.2)
(6)	Miami-Comparison Difference	−2.0 (1.9)	−3.0 (2.0)	−1.0 (2.8)

[a] Notes: Adapted from Card (1990, Tables 3 and 6). Standard errors are shown in parentheses.

Table 4 illustrates DD estimation of the effect of Boatlift immigrants on unemployment rates, separately for whites and blacks. The first column reports unemployment rates in 1979, the second column reports unemployment rates in 1981, and the third column reports the 1981–1979 difference. The rows give numbers for Miami, the comparison cities, and the difference between them. For example, between 1981 and 1979, the unemployment rate for Blacks in Miami rose by about 1.3%, though this change is not significant. Unemployment rates in the comparisons cities rose even more, by 2.3%. The difference in these two changes, −1.0%, is a DD estimate of the effect of the Mariel immigrants on the unemployment rate of Blacks in Miami. In this case, the estimated effect on the unemployment rate is actually negative, though not significantly different from zero.

The rationale for this double-differencing strategy can be explained in terms of restrictions on the conditional mean function for potential outcomes in the absence of immigration. As in the union example, let Y_{0i} be i's employment status in the absence of immigration and let Y_{1i} be i's employment status if the Mariel immigrants come to i's city. The unemployment rate in city c in year t is $\mathrm{E}[Y_{0i} \mid c, t]$, with no immigration wave, and $\mathrm{E}[Y_{1i} \mid c, t]$ if there is an immigration wave. In practice, we know that the Mariel immigration happened in Miami in 1980, so that the only values of $\mathrm{E}[Y_{1i} \mid c, t]$ we get to see are for $c =$ Miami and $t > 1980$. The Mariel Boatlift study uses the comparison cities to estimate the counterfactual average, $\mathrm{E}[Y_{0i} \mid c =$ Miami, $t > 1980]$, i.e., what the unemployment rate in Miami would have been if the Mariel immigrants had not come.

The DD method identifies causal effects by restricting the conditional mean function $\mathrm{E}[Y_{0i} \mid c, t]$ in a particular way. Specifically, suppose that

$$\mathrm{E}[Y_{0i} \mid c, t] = \beta_t + \gamma_c, \tag{18}$$

that is, in the absence of immigration, unemployment rates can be written as the sum of a year effect that is common to cities and a city effect that is fixed over time. The additive model pertains to $E[Y_{0i} \mid c, t]$ instead of Y_{0i} directly because the latter is a zero/one variable. Suppose also that the effect of the Mariel immigration is simply to add a constant to $E[Y_{0i} \mid c, t]$, so that

$$E[Y_{1i} \mid c, t] = E[Y_{0i} \mid c, t] + \delta. \tag{19}$$

This means the employment status of individuals living in Miami and the comparison cities in 1979 and 1981 can be written as

$$Y_i = \beta_t + \gamma_c + \delta M_i + \varepsilon_i, \tag{20}$$

where $E[\varepsilon_i \mid c, t] = 0$ and M_i is a dummy variable that equals 1 if i was exposed to the Mariel immigration by living in Miami after 1980. Differencing unemployment rates across cities and years gives

$$\{E[Y_i \mid c = \text{Miami}, t = 1981] - E[Y_i \mid c = \text{Comparison}, t = 1981]\}$$

$$-\{E[Y_i \mid c = \text{Miami}, t = 1979] - E[Y_i \mid c = \text{Comparison}, t = 1979]\} = \delta. \tag{21}$$

Note that M_i in Eq. (20) is an interaction term equal to the product of a dummy indicating observations after 1980 and a dummy indicating residence in Miami. The DD estimate can therefore also be computed in a regression of stacked micro data for cities and years. The regressors consist of dummies for years, dummies for cities, and M_i. Similarly, a regression-adjusted version of the DD estimator adds a vector of individual characteristics, X_i to Eq. (20):

$$Y_i = X_i'\beta_0 + \beta_t + \gamma_c + \delta M_i + \varepsilon_i,$$

where β_0 is now a vector of coefficients that includes a constant. Controlling for X_i changes the estimate of δ only if M_i are X_i are correlated, conditional on city and year main-effects. (In practice, δ might be allowed to differ for different post-treatment years.)

DD pitfalls. Like any other identification strategy, DD is not guaranteed to identify the causal effect of interest. Meyer (1995) and Campbell (1969) outline a range of threats to the causal interpretation of DD estimates. The key identifying assumption is clearly that interaction terms are zero in the absence of the intervention. In fact, it is easy to imagine that unemployment rates evolve differently across cities regardless of shocks like the Mariel immigration. One way to test this is to compare trends in outcomes before or after the event of interest. As noted above, the comparison cities in this case were chosen partly on the basis of Fig. 1, which shows that the comparison cities exhibited a pattern of economic growth similar to that in Miami. Identification of causal effects using city/year comparisons clearly turns on the assumption that the two sets of cities would have had the same employment trends had the boatlift not occurred. We introduce some new evidence on this question in Section 2.4.

2.2.3. Instrumental variables

Identification strategies based on instrumental variables can be thought of as a scheme for using exogenous field variation to approximate randomized trials. Again, we illustrate with an example where there is an underlying causal relationship, in this case the effect of Vietnam-era military service on the earnings of veterans later in life. In the 1960s and early 1970s, young men were at risk of being drafted for military service. Policy makers, veterans groups, and economists have long been interested in what the consequences of this military service were for the men involved. A belief that military service is a burden helped to mobilize support for a range of veterans' programs and for ending the draft in 1973 (see, e.g., Taussig, 1974). Concerns about fairness also led to the institution of a draft lottery in 1970 that was used to determine priority for conscription in cohorts of 19-year-olds. This lottery was used by Hearst et al. (1986) to estimate the effects of military service on civilian mortality and by Angrist (1990) to construct IV estimates of the effects of military service on civilian earnings.

As in the union problem, the causal relationship of interest is based on the notion that there are two potential outcomes, Y_{0i}, denoting what someone from the Vietnam-era cohort would earn if they did not serve in the military and Y_{1i}, denoting earnings as a veteran. Again, using a constant-effects model for potential outcomes, we can write

$$Y_{0i} = \beta_0 + \eta_i, \qquad Y_{1i} = Y_{0i} + \delta, \tag{22}$$

where $\beta_0 \equiv \mathrm{E}[Y_{0i}]$. The constant effect δ is the parameter of interest. IV estimates have a causal interpretation under weaker assumptions than this, but we postpone a discussion of this point until Section 2.3. As in the union and schooling problems, η_i is the random part of potential outcomes, but at this point there are no observed covariates in the model for Y_{0i}. Using D_i to indicate veteran status, the causal relationship between veteran status and earnings can be written

$$Y_i = \beta_0 + D_i\delta + \eta_i. \tag{23}$$

Also as in the union and schooling problems, there is a concern that since D_i is not randomly assigned, a comparison of all veterans to all non-veterans would not identify δ. Suppose, for example, that individuals with low civilian earnings potential are more likely to serve in the military, either because they want to or because they are less adept at obtaining deferments. Then the regression coefficient in (23), which is also the difference in means by veteran status, is biased downwards:

$$\mathrm{E}[Y_i \mid D_i = 1] - \mathrm{E}[Y_i \mid D_i = 0] = \delta + \{\mathrm{E}[\eta_i \mid D_i = 1] - \mathrm{E}[\eta_i \mid D_i = 0]\} < \delta. \tag{24}$$

IV methods can eliminate this sort of bias if the researcher has access to an instrumental variable Z_i that is correlated with D_i, but otherwise independent of potential outcomes. A natural instrument is draft-eligibility status, since this was determined by a lottery over birthdays. In particular, in each year from 1970 to 1972, random sequence numbers (RSNs) were randomly assigned to each birth date in cohorts of 19-year-olds. Men with lottery numbers below an eligibility ceiling were eligible for the draft, while men with

Table 5
IV estimates of the effects of military service on white men[a]

Earnings year	Earnings		Veteran status		Wald estimate of veteran effect
	Mean (1)	Eligibility effect (2)	Mean (3)	Eligibility effect (4)	(5)
A. Men born 1950					
1981	16461	−435.8 (210.5)	0.267	0.159 (0.040)	−2741 (1324)
1970	2758	−233.8 (39.7)			−1470 (250)
1969	2299	−2.0 (34.5)			
B. Men born 1951					
1981	16049	−358.3 (203.6)	0.197	0.136 (0.043)	−2635 (1497)
1971	2947	−298.2 (41.7)			−2193 (307)
1970	2379	−44.8 (36.7)			
C. Men born 1953 (no one drafted)					
1981	14762	34.3 (199.0)	0.130	0.043 (0.037)	No first stage
1972	3989	−56.5 (54.8)			
1971	2803	2.1 (42.9)			

[a] Note: Adapted from Angrist (1990, Tables 2 and 3), and unpublished author tabulations. Standard errors are shown in parentheses. Earnings data are from Social Security administrative records. Figures are in nominal dollars. Veteran status data are from the Survey of Program Participation. There are about 13,500 observations with earnings in each cohort.

numbers above the ceiling could not be drafted. In practice, many draft-eligible men were still exempted from service for health or other reasons, while many men who were draft-exempt nevertheless volunteered for service. So veteran status was not completely determined by randomized draft-eligibility; eligibility and veteran status are merely correlated.

For white men who were at risk of being drafted in the 1970–1971 draft lotteries, draft-eligibility is clearly associated with lower earnings in years after the lottery. This can be seen in Table 5, which reports the effect of randomized draft-eligibility status on Social Security earnings in column (2). Column (1) shows average annual earnings for purposes of comparison. These data are the FICA-taxable earnings of men with earnings covered by OASDI (for details see the appendix to Angrist (1990)). For men born in 1950, there are significant negative effects of eligibility status on earnings in 1970, when these men were being drafted, and in 1981, 10 years later. In contrast, there is no evidence of an association between eligibility status and earnings in 1969, the year the lottery drawing for men born in 1950 was held but before anyone born in 1950 was actually drafted. Similarly, for men born in 1951, there are large negative eligibility effects in 1971 and 1981, but no evidence of an effect in 1970, before anyone born in 1951 was actually drafted. The timing of these effects suggests that the negative association between draft-eligibility status and earnings is caused by the military service of draft-eligible men.

Because eligibility status was randomly assigned, the claim that the estimates in column

(2) represent the effect of *draft-eligibility* on earnings seems uncontroversial. How do we go from the effect of draft-eligibility to the effect of veteran status? The identifying assumption in this case is that Z_i is independent of potential earnings, which in this case means that Z_i is uncorrelated with η_i. It follows immediately that $\delta = C(Y_i, Z_i)/C(D_i, Z_i)$. The intuition here is that only part of the variation in D_i – the part that is associated with Z_i – is used to identify the parameter of interest (δ). Because Z_i is a binary variable, we also have

$$\delta = \{E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0]\}/\{E[D \mid Z_i = 1] - E[D \mid Z_i = 0]\}. \tag{25}$$

The sample analog of (25) is the Wald (1940) estimator that was originally applied to measurement error problems.[15] Note that we could have arrived at (25) directly, i.e., without reference to the $C(Y_i, Z_i)/C(D_i, Z_i)$ formula, because the independence of Z_i and potential outcomes implies $E[\eta_i \mid Z_i] = 0$. In this case, the Wald estimator is simply the difference in mean earnings between draft-eligible and ineligible men, divided by the difference in the probability of serving in the military between draft-eligible and ineligible men.

The only information required to go from draft-eligibility effects to veteran-status effects is the denominator of the Wald estimator, which is the effect of draft-eligibility on the probability of serving in the military. This information, which comes from the Survey of Income and Program Participation (SIPP), appears in column (4) of Table 5.[16] For earnings in 1981, long after most Vietnam-era servicemen were discharged from the military, the Wald estimates of the effect of military service amount to about 16% of earnings. Effects for men while in the service are much larger (in percentage terms), which is not surprising since military pay during the conscription era was extremely low.

An important feature of the Wald/IV estimator is that the identifying assumptions are easy to assess and interpret. The basic claim justifying a causal interpretation of the estimator is that the only reason why $E[Y_i \mid Z_i]$ varies with Z_i is because $E[D_i \mid Z_i]$ varies with Z_i. A simple way to check this is to look for an association between Z_i and personal characteristics that should not be affected by D_i, such as age, race, sex, or any other characteristic that was determined before D_i was determined. Another useful check is to look for an association between the instrument and outcomes in samples where there is no reason for such a relationship. If it really is true that the only reason why draft-eligibility affects earnings is veteran status, then in samples where eligibility status is unrelated to veteran status, draft-eligibility effects on earnings should be zero. This idea is illustrated in section C of Table 5, which reports estimates for men born in 1953. Although there was a lottery drawing which assigned RSNs to the 1953 cohort in February of 1972, no one born in 1953 was actually drafted (the draft officially ended in July 1973). This is reflected in

[15] The relationship between IV with binary instruments and Wald estimators was first noted by Durbin (1954).

[16] In this case, the denominator of the Wald estimates does not come from the same data set as the numerator since the Social Security administration has no information on veteran status. As long as the information used to estimate the numerator and denominator are representative of the same population, the resulting two-sample estimate will be consistent. The econometrics behind this two-sample approach to IV are discussed briefly in Section 3.4.

the insignificant first-stage relationship between veteran status and draft-eligibility for men born in 1953 (defined using the 1952 RSN cutoff of 95). In fact, there is no significant relationship between Y_i and Z_i for this cohort as well. Evidence of a relationship between Z_i and Y_i would cast doubt on the claim that the only reason for draft-eligibility effects is the military service of the men who were draft-eligible. We discuss other specification checks of this type in Section 2.4.

So far the discussion of IV has allowed for only three variables: the outcome, the endogenous regressor, and the instrument. In many cases, the assumption that $E[Z_i\eta_i] = 0$ is more plausible after controlling for a vector of covariates, X_i. Decomposing the random part of potential outcomes in (22) into a linear function of k control variables and an error term so that $\eta_i = X_i'\beta + \varepsilon_i$ as before, the resulting estimating equation is

$$Y_i = X_i'\beta + D_i\delta + \varepsilon_i. \tag{26}$$

Note that since ε_i is defined as the residual from a regression of η_i on X_i, it is uncorrelated with X_i by construction. In contrast with δ, which has a causal interpretation., the coefficient vector β is not meant to capture the causal effect of the X-variables. As in the discussion of regression, we find it useful to distinguish between control variables and causing variables when using instrumental variables.

Equations like (26) are typically estimated using 2SLS, i.e., by substituting the fitted values from a first-stage regression of D_i on X_i and Z_i. In some applications, more than one instrument is available to estimate the single causal effect, δ. 2SLS accommodates this situation by including all the instruments in the first-stage equation. The combination of multiple instruments to produce a single estimate makes the most sense in a constant-coefficients framework. The assumptions of instrument validity and constant coefficients can also be tested in this case (see, e.g., Hansen, 1982; Newey, 1985). In a more general setting with heterogeneous potential outcomes, different instruments estimate different weighted averages of the difference $Y_{1i} - Y_{0i}$ (Imbens and Angrist, 1994). We return to this point in Section 2.3.

IV pitfalls. The most important IV pitfall is the validity of instruments, i.e., the possibility that η_i and Z_i are correlated. Suppose, for example, that Z_i is related to the vector of control variables, X_i, and we do not account for this in the estimation. The Wald/IV estimator in that case has probability limit

$$\delta + \beta'\{E[X_i \mid Z_i = 1] - E[X_i \mid Z_i = 0]\}/\{E[D_i \mid Z_i = 1] - E[D_i \mid Z_i = 0]\}.$$

This is a version of the omitted-variables bias formula for IV. The formula captures the fact that "a little omitted variables bias can go a long way" in an IV setting, because the association between X_i and Z_i gets multiplied by $\{E[D \mid Z = 1] - E[D \mid Z = 0]\}^{-1}$. In the draft lottery case, for example, any draft-eligibility effects on omitted variables get multiplied by about $1/0.15 \approx 6.7$.

A second important point about bias in instrumental variables estimates is that random assignment alone does not guarantee a valid instrument. Suppose, for example, that in

addition to being more likely to serve in the military, men with low draft-lottery numbers were more likely to stay in college so as to extend a draft deferment. This fact will create a relationship between potential earnings and Z_i even for non-veterans, in which case IV yields biased estimates of the causal effect of veteran status. Random assignment of Z_i does not rule out this sort of bias since draft-eligibility can in principle have consequences in addition to influencing the probability of being a veteran. In other words, while the randomization of Z_i ensures that the reduced-form relationship between Y_i and Z_i represents the causal effect of draft eligibility on earnings, it does not guarantee that the only reason for this relationship is D_i. The distinction between the assumed random assignment of an instrument and the assumption that a single causal mechanism explains effects on outcomes is discussed in greater detail by Angrist et al. (1996).

Finally, the use of 2SLS to combine many different instruments can lead to finite-sample bias. The standard inference framework for 2SLS uses asymptotic theory, i.e., inference is based on approximations that are increasingly accurate as sample sizes grow. Typically, inferences about OLS coefficient estimates also use asymptotic theory since the relevant finite-sample theory assumes normally distributed errors. A key difference between IV and OLS estimators, however, is that even without normality OLS provides an unbiased estimate of population regression coefficients (provided the regression function is linear; see, e.g., Goldberger, 1991, Chapter 13). In contrast, IV estimators are consistent but not unbiased. This means that under repeated sampling with a fixed sample size, IV estimates may systematically deviate from the corresponding population parameter.[17] Moreover, this bias tends to pull IV estimates towards the corresponding OLS estimates, giving a misleading impression of similarity between the two sets of estimates (see, e.g., Sawa, 1969).

How bad is the finite-sample bias of an IV estimate likely to be? In practice, this largely turns on the number of instruments relative to the sample size, and the strength of the first-stage relationship. Other things equal, more instruments, smaller samples, and weaker instruments each mean more bias (see, e.g., Buse, 1992). The fact that IV estimates can be noticeably biased even with very large datasets was highlighted by Bound et al. (1995), which focuses on Angrist and Krueger's (1991) compulsory schooling study. This study uses hundreds of thousands of observations from Census data to implement an instrumental variables strategy for estimating the returns to schooling. The instruments are quarter-of-birth dummies since children born earlier in the year enter school at an older age and are therefore allowed to drop out of school (typically on their 16th birthday) after having completed less schooling. Some of the 2SLS estimates in Angrist and Krueger (1991) use many quarter-of-birth/state-of-birth interaction terms in addition to quarter-of-birth main effects as instruments. Since the underlying first-stage relationship in these models is not very strong, there is potential for substantial bias towards the OLS estimates in these specifications.

[17] A similar problem arises with Generalized Method of Moments estimation of models for covariance structures (see Altonji and Segal, 1996).

Bound et al. (1995) discuss the question of how strong a first-stage relationship has to be in order to minimize the potential for bias. They suggest using the F-statistic for the joint significance of the excluded instruments in the first-stage equation as a diagnostic. This is clearly sensible, since, if the instruments are so weak that the relationship between instruments and endogenous regressors cannot be detected with a reasonably high level of confidence, then the instruments should probably be abandoned. On the other hand, Hall et al. (1996) point out that this sort of selection procedure also has the potential to induce a bias from pre-testing.

A simple alternative (or complement) to screening on the first-stage F is to use estimators that are approximately unbiased. One such estimator is Limited Information Likelihood (LIML), which has no integral moments but is nevertheless median-unbiased. This means that the sampling distribution is centered at the population parameter.[18] In fact, any just-identified 2SLS estimator is also median-unbiased since 2SLS and LIML are identical for just-identified models. The class of median-unbiased instrumental variables estimators therefore includes the Wald estimator discussed in the previous section. Other approximately unbiased estimators are based on procedures that estimate the first-stage and second-stage relationship in separate datasets. This includes Two-Sample and Split-Sample IV (Angrist and Krueger, 1992, 1995), and an IV estimator that uses a set of leave-one-out first-stage estimates called Jackknife Instrumental Variables (Angrist et al., 1998).[19] An earlier literature discussed combination estimators that are approximately unbiased (see, e.g., Sawa, 1973). Recently, Chamberlain and Imbens (1996) introduced a Bayesian IV estimator that also avoids bias.

A final and related point is that the reduced-form OLS regression of the dependent variable on exogenous covariates and instruments is unbiased in a sample of any size, regardless of the power of the instrument (assuming the reduced form is linear). This is important because the reduced form effects of the instrument on the dependent variable are proportional to the coefficient on the endogenous regressor in the equation of interest. The existence of a causal relationship between the endogenous regressor and dependent variable can therefore be gauged through the reduced form without fear of finite-sample bias even if the instruments are weak.

2.2.4. *Regression-discontinuity designs*

The Latin motto Marshall placed on the title page of his *Principles of Economics* (Marshall, 1890) is, *"Natura non facit saltum,"* which means: "Nature does not make

[18] Anderson et al. (1982, p. 1026) report this in a Monte Carlo study: "To summarize, the most important conclusion from the study of LIML and 2SLS estimators is that the 2SLS estimator can be badly biased and in that sense its use is risky. The LIML estimator, on the other hand, has a little more variability with a slight chance of extreme values, but its distribution is centered at the parameter value." Similar Monte Carlo results and a variety of analytic justifications for the approximate unbiasedness of LIML appear in Bekker (1994), Donald and Newey (1997), Staiger and Stock (1997), and Angrist et al. (1998).

[19] A SAS program that computes Split-Sample and Jackknife IV is available at http://www.wws.princeton.edu/faculty/krueger.html.

jumps." Marshall argues that most economic behavior evolves gradually enough to be modeled or explained. The notion that human behavior is typically orderly or smooth is at the heart of a research strategy called the regression-discontinuity (RD) design. RD methods use some sort of parametric or semi-parametric model to control for smooth or gradually evolving trends, inferring causality when the variable of interest changes abruptly for non-behavioral or arbitrary reasons. There are a number of ways to implement this idea in practice. We focus here on an approach that can viewed as a hybrid regression-control/IV identification strategy. This is distinct from conventional IV strategies because the instruments are derived explicitly from non-linearities or discontinuities in the relationship between the regressor of interest and a control variable. Recent applications of the RD idea include van der Klauuw's (1996) study of financial aid awards; Angrist and Lavy's (1998) study of class size; and Hahn et al.'s (1998) study of anti-discrimination laws.

The RD idea originated with Campbell (1969), who discussed the (theoretical) problem of how to identify the causal effect of a treatment that is assigned as a deterministic function of an observed covariate which is also related to the outcomes of interest. Campbell used the example of estimating the effect of National Merit scholarships on applicants' later academic achievement. He argued that if there is a threshold value of past achievement that determines whether an award is made, then one can control for any smooth function of past achievement and still estimate the effect of the award at the point of discontinuity. This is done by matching discontinuities or non-linearities in the relationship between outcomes and past achievement to discontinuities or non-linearities in the relationship between awards and past achievement.[20] van der Klauuw (1996) pointed out the link between Campbell's suggestion and IV, and used this idea to estimate the effect of financial aid awards on college enrollment.[21]

Angrist and Lavy (1998) used RD to estimate the effects of class size on pupil test scores in Israeli public schools, where class size is officially capped at 40. They refer to the cap of 40 as "Maimonides' Rule," after the 12th Century Talmudic scholar Maimonides, who first proposed it. According to Maimonides' Rule, class size increases one-for-one with enrollment until 40 pupils are enrolled, but when 41 students are enrolled, there will be a sharp drop in class size, to an average of 20.5 pupils. Similarly, when 80 pupils are enrolled, the average class size will again be 40, but when 81 pupils are enrolled the average class size drops to 27. Thus, Maimonides' Rule generates discontinuities in the relationship between grade enrollment and average class size at integer multiples of 40.

The class size function derived from Maimonides' Rule can be stated formally as

[20] Goldberger (1972) discusses a similar idea in the context of compensatory education programs.

[21] Campbell's (1969) discussion of RD focused mostly on what he called a "sharp design", where the regressor of interest is a discontinuous but deterministic function of another variable. In the sharp design there is no need to instrument – the regressor of interest is entered directly. This is in contrast with what Campbell called a "fuzzy design", where the function is not deterministic. Campbell did not propose an estimator for the fuzzy design, though his student Trochim (1984) developed an IV-like procedure for that case. The discussion here covers the fuzzy design only since the sharp design can be viewed as a special case.

follows. Let b_s denote beginning-of-the-year enrollment in school s in a given grade, and let z_s denote the size assigned to classes in school s, as predicted by applying Maimonides' Rule to that grade. Assuming cohorts are divided into classes of equal size, the predicted class size for all classes in the grade is

$$z_s = b_s/(\text{int}((b_s - 1)/40) + 1).$$

This function is plotted in Fig. 2A for the population of Israeli fifth graders in 1991, along with actual fifth grade class sizes. The x-axis shows September enrollment and the y-axis shows either predicted class size or the average actual class size in all schools with that enrollment. Maimonides' Rule does not predict actual class size perfectly because other factors affect class size as well, but average class sizes clearly display a sawtooth pattern induced by the Rule.

In addition to exhibiting a strong association with average class size, Maimonides' Rule is also correlated with average test scores. This is shown in Fig. 2B, which plots average reading test scores and average values of z_s by enrollment size, in enrollment intervals of 10. The figure shows that test scores are generally higher in schools with larger enrollments and, therefore, larger predicted class sizes. Most importantly, however, average scores by enrollment size exhibit a sawtooth pattern that is, at least in part, the mirror image of the class size function. This is especially clear in Fig. 2C, which plots average scores by enrollment after running auxiliary regressions to remove a linear trend in enrollment and the effects of pupils' socioeconomic background.[22] The up and down pattern in the conditional expectation of test scores given enrollment probably reflects the causal effect of changes in class size that are induced by exogenous changes in enrollment. This interpretation is plausible because Maimonides' Rule is known to have this pattern, while it seems likely that other mechanisms linking enrollment and test scores will be smoother.

Fig. 2B makes it clear that Maimonides' Rule is not a valid instrument for class size without controlling for enrollment because predicted class size increases with enrollment and test scores increase with enrollment. The RD idea is to use the discontinuities (jumps) in predicted class size to estimate the effect of interest while controlling for smooth enrollment effects. Angrist and Lavy implement this by using z_s as an instrument while controlling for smooth effects of enrollment using parametric enrollment trends. Consider a causal model that links the score of pupil i in school s with class size and school characteristics:

$$y_{is} = X_s'\beta + n_{is}\delta + \varepsilon_{is}, \tag{27}$$

where n_{is} is the size of i's class, and X_s is a vector of school characteristics, including functions of grade enrollment, b_s. As before, we imagine that this function tells us what test

[22] The figure plots the residuals from regressions of y_{is} and z_s on b_s and the proportion of low-income pupils in the school.

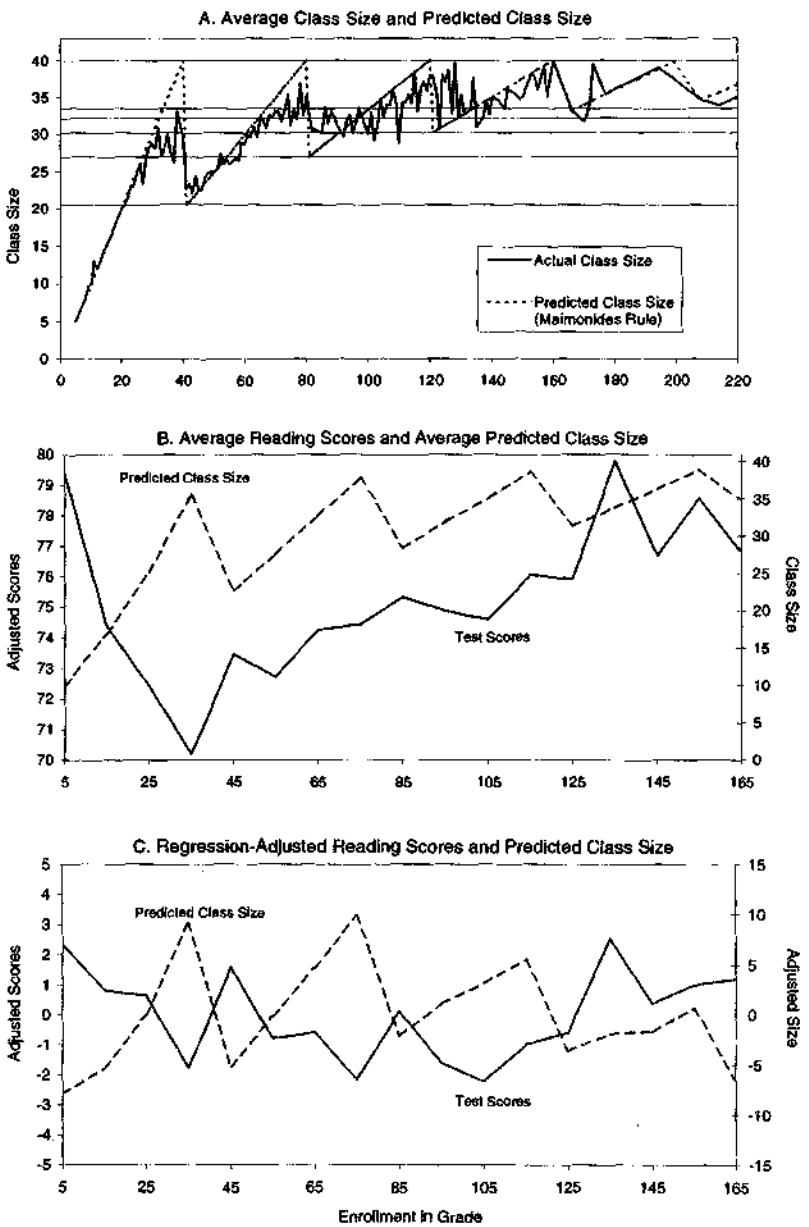

Fig. 2. Illustration of regression-discontinuity method for estimating the effect of class size on pupil's test scores. Data are from Angrist and Lavy (1998).

scores would be if class size were manipulated to be other than the observed size, n_{is}. The first-stage equation for 2SLS estimation of (27) is

$$n_{is} = X_s'\pi_0 + z_s\pi_1 + v_{is}. \tag{28}$$

A simple example is a model that includes b_s linearly to control for enrollment effects not attributable to changing class size, along with a regressor measuring the proportion of low-income students in the school.[23] The resulting 2SLS estimate of δ in standard deviation units is -0.037 (with a standard error of 0.009), meaning just over a one-third standard deviation decline in test scores for a 10 pupil increase in class size.

Since RD is an IV estimator, we do not have a separate section for pitfalls. As before, the most important issue is instrument validity and the choice of control variables. The choice of controls is even more important in RD than conventional IV, however, since the instrument is actually a function of one of the control variables. In the Angrist and Lavy application, for example, identification of δ clearly turns on the ability to distinguish z_s from X_s since z_s does not vary within schools. This suggests that RD depends more on functional form assumptions than other IV procedures, although Hahn et al. (1998) consider ways to weaken this dependence.

2.3. Consequences of heterogeneity and non-linearity

The discussion so far involves a highly stylized description of the world, wherein causal effects are the same for everyone, and, if the causing variable takes on more than two values, the effects are linear. Although some economic models can be used to justify these assumptions, there is no reason to believe they are true in general. On the other hand, these strong assumptions provide a useful starting place because they may provide a good approximation of reality, and because they focus attention on basic causality issues.

The cost of these simplifying assumptions is that they gloss over the fact that even when a set of estimates has a causal interpretation, they are generated by variation for a particular group of individuals over a limited range of variation in the causing variable. There is a tradition in Psychology of distinguishing between the question of *internal validity*, i.e., whether an empirical relationship has a causal interpretation in the setting where it is observed, and the question of *external validity*, i.e., whether a set of internally valid estimates has predictive value for groups or values of the response variable other than those observed in a given study.[24] Constant-coefficient and linear models make it harder to discuss the two types of validity separately, since external validity is automatic in a constant-coefficients-linear setting. For example, the constant-effects model says that the economic consequences of military service are the same for high-school dropouts and college graduates. Similarly, the linear model says the economic value of a year of

[23] In practice, Angrist and Lavy estimated (27) and (28) using class-level averages and not micro data.

[24] See, e.g., Campbell and Stanley (1963) and Meyer (1995).

schooling is the same whether the year is second grade or the last year of college. We therefore discuss the interpretation of traditional estimators when constant-effects and linearity assumptions are relaxed.

2.3.1. Regression and the conditional expectation function

Returning to the schooling example of Section 2.2.1, the causal relationship of interest is $f_i(S)$, which describes the effect of schooling on earnings. In the absence of any further assumptions, the average causal response function is $E[f_i(S)]$, with average derivative $E[f_i'(S)]$. Earlier, we assumed $f_i'(S)$ is equal to a constant, ρ, in which case averaging is not needed. In practice, however, the derivative may be heterogeneous; that is, it may vary with i or with i's characteristics, X_i. In economics, models for heterogenous treatment effects are commonly called "random coefficient" models (see, e.g., Björklund and Moffitt, 1987 or Heckman and Robb, 1985 for discussions of such models). The derivative also might be non-constant (i.e., vary with S). In either case, it makes sense to focus on the average response function or its average derivative. The principal statistical tool for doing this is the Conditional Expectation Function (CEF) of Y_i given S_i, i.e., $E[Y_i \mid S_i = S]$ or $E[Y_i \mid X_i, S_i = S]$, viewed as a function of S.

To see the connection between the CEF and the average causal response, consider first the difference in average earnings between people with S years of schooling and people with $S - 1$ years of schooling:

$$E[Y_i \mid S_i = S] - E[Y_i \mid S_i = S - 1] = E[f_i(S) - f_i(S - 1) \mid S_i = S]$$

$$+ \{E[f_i(S - 1) \mid S_i = S] - E[f_i(S - 1) \mid S_i = S - 1]\}.$$

The first term in this decomposition is the average causal effect of going from $S - 1$ to S years of schooling for those who actually have S years of education. The counterfactual average $E[f_i(S - 1) \mid S_i = S]$ is never observed, however. The second term reflects the fact that the average earnings of those with $S - 1$ years of schooling do not necessarily provide a good answer to the "what if" question for those with S years of schooling. This term is the counterpart of regression-style "omitted variables bias" for this more general model.

In this setting, the selection-on-observables assumption asserts that conditioning on a set of observed characteristics, X_i, serves to eliminate the omitted variables bias in naive comparisons. That is,

$$E[f_i(S - 1) \mid X_i, S_i = S] = E[f_i(S - 1) \mid X_i, S_i = S - 1] \qquad \text{for all } S, \tag{29}$$

so that conditional on X, the CEF and average causal response function are the same:

$$E[Y_i \mid X_i, S_i = S] = E[f_i(S) \mid X_i].$$

In this case, the conditional-on-X comparison does estimate the causal effect of schooling:

$$E[Y_i \mid X_i, S_i = S] - E[Y_i \mid X_i, S_i = S - 1] = E[f_i(S) - f_i(S - 1) \mid X_i].$$

This is analogous to the notion that adding X_i to a regression eliminates omitted variables bias in OLS estimates of the returns to schooling.

The preceding discussion provides sufficient conditions for the CEF to have a causal interpretation. We next consider the relationship between regression parameters and the CEF. One interpretation of regression is that the population OLS slope vector provides a minimum mean squared error (MMSE) linear approximation to the CEF. This feature of regression is discussed in Goldberger's (1991) econometrics text (see especially Section 5.5).[25] A related property is the fact that regression coefficients have an "average derivative" interpretation. In multivariate regression models, however, this interpretation is complicated by the fact that the OLS slope vector is actually matrix-weighted average of the gradient of the CEF. Matrix-weighted averages are difficult to interpret except in special cases (see Chamberlain and Leamer, 1976).[26]

One interesting special case where the OLS slope vector can be readily interpreted is when S_i is the single regressor of interest and the CEF of this regressor given all other regressors is linear, so that

$$\mathrm{E}[S_i \mid X_i] = X_i'\pi, \tag{30}$$

where π is a conformable vector of coefficients. This assumption is satisfied in the schooling regression, for example, in a model where all X-variables are discrete and the parameterization allows a separate effect for each possible value of X_i. This is not unrealistic in applications with large datasets; see, for example, Angrist and Krueger (1991) and Angrist (1998). In this case, the population regression coefficient from a regression of Y_i on X_i and S_i can be written

$$\begin{aligned}\rho_x &= \mathrm{E}[(S_i - \mathrm{E}[S_i \mid X_i])Y_i]/\mathrm{E}[(S_i - \mathrm{E}[S_i \mid X_i])S_i] \\ &= \mathrm{E}[(S_i - \mathrm{E}[S_i \mid X_i])\mathrm{E}[Y \mid X_i, S_i]]/\mathrm{E}[(S_i - \mathrm{E}[S_i \mid X_i)S_i],\end{aligned} \tag{31}$$

which is derived by iterating expectations over X_i and S_i.

Maintaining assumption (30), i.e., that $\mathrm{E}[S_i \mid X_i]$ is linear, first consider the case where $\mathrm{E}[Y_i \mid X_i, S_i]$ is linear in S_i but not X_i. Then we can write

$$\rho_X \equiv \mathrm{E}[Y_i \mid X_i, S_i = S] - \mathrm{E}[Y_i \mid X_i, S_i = S - 1],$$

for all S, which means

[25] Proof that OLS gives a MMSE linear approximation to the CEF: The vector of population regression coefficients for regressor vector W_i solves $\min_b \mathrm{E}(Y_i - W_i'b)^2$. But $(Y_i - W_i'b)^2 = [(Y_i - \mathrm{E}[Y_i \mid W_i]) + (\mathrm{E}[Y_i \mid W_i] - W_i'b)]^2$ and $\mathrm{E}[(Y_i - \mathrm{E}[Y_i \mid W_i])\,(\mathrm{E}[Y_i \mid W_i] - W_i'b)] = 0$, so $\min_b \mathrm{E}([Y_i \mid W_i] - W_i'b)]^2$ has the same solution.

[26] The population slope vector is $\mathrm{E}[W_iW_i']^{-1}\mathrm{E}[W_iY_i] = \mathrm{E}[W_iW_i']^{-1}\mathrm{E}[W_i\mathrm{E}(Y_i \mid W_i)]$. Assume $\mathrm{E}(W_i) = 0$ so these are the non-intercept coefficients. Linearizing the CEF, we have $\mathrm{E}(Y_i \mid W_i) = \mathrm{E}(Y_i \mid W_i = 0) + W_i'\nabla\mathrm{E}(Y_i \mid \tilde{w}_i)$, where $\nabla\mathrm{E}(Y_i \mid \tilde{w}_i)$ is the gradient of the conditional expectation function, and $\tilde{w}_i$ is a random vector that lies between W_i and zero. So the slope vector is $\mathrm{E}[W_iW_i']^{-1}\mathrm{E}[(W_iW_i')\nabla\mathrm{E}(Y_i \mid \tilde{w}_i)]$, which is a matrix-weighted average of the gradient with weights (W_iW_i').

$$E[Y_i \mid X_i, S_i] = E[Y_i \mid X_i, S_i = 0] + S_i\rho_X. \tag{32}$$

In other words, the CEF is linear in schooling, but the schooling coefficient is not constant and depends on X_i.

Substituting (32) into (31), we have

$$\rho_r = E[(S_i - E[S_i \mid X_i])^2\rho_X]/E[(S_i - E[S_i \mid X_i])^2] = E[\sigma_S^2(X_i)\rho_X]/E[\sigma_S^2(X_i)], \tag{33}$$

where $\sigma_S^2(X_i) \equiv E[S_i - E[S_i|X_i])^2|X_i]$ is the variance of S_i given X_i. So in this case, regression provides a variance-weighted average of the slope at each X_i. Values of X_i that get the most weight are those where the conditional variance of schooling is largest.

What if the CEF of Y_i varies with both X_i and S_i? Let

$$\rho_{SX} \equiv E[Y_i \mid X_i, S_i = S] - E[Y_i \mid X_i, S_i = S - 1],$$

where the ρ_{SX} notation reflects variation with both S and X_i. Then the coefficient on S_i in a regression of Y_i on X_i and S_i can be written

$$\rho_r = E\left[\sum_{S=1}^{\bar{S}} \rho_{SX}\mu_{SX}\right]E\left[\sum_{S=1}^{\bar{S}} \mu_{SX}\right]^{-1}, \tag{34}$$

where

$$\mu_{SX} = (E[S_i \mid X_i, S_i \geq S] - E[S_i \mid X_i, S_i < S])P[S_i \geq S \mid X_i](1 - P[S_i \geq S \mid X_i]) \geq 0.$$

and S takes on values in the set $\{0, 1, \ldots, \bar{s}\}$. This result, which is proved in Appendix A, is a generalization of the formula for bivariate regression coefficients given by Yitzhaki (1996).[27]

The weighting formula in (34) has a sum and an expectation. The sum averages ρ_{SX} for all schooling increments, given a particular value of X_i (this averaging matters if the CEF is non-linear). The expectation then averages this sum in the distribution of X_i (this averaging matters if the response function is heterogeneous). The formula for the weights, μ_{SX}, can be used to characterize the OLS slope vector. First, for any particular X_i, weight is given to ρ_{SX} for each S in proportion to the change in the conditional mean of S_i, as S_i falls above or below S. More weight is also given to points in the domain of $f_i(S)$ that are close to the conditional median of S_i given X_i since this is where $P[S_i \geq S \mid X_i](1 - P[S_i \geq S \mid X_i])$ is maximized. Second, as in the linear case discussed above, weight is also given in proportion to conditional variance of S_i given X_i, except now this variance is defined separately for each S using dummies for the event that $S_i \geq S$. Note also that the OLS estimate contains no information about the returns to schooling for values of X_i where

[27] Yitzhaki gives examples and describes the OLS weighting function for a model with a single continuously distributed regressor in detail. For Normally distributed regressors, the weighting function is the Normal density function, so that OLS provides a density-weighted average of the sort discussed by Powell et al. (1989). For an alternative non-parametric interpretation of OLS coefficients see Stoker (1986).

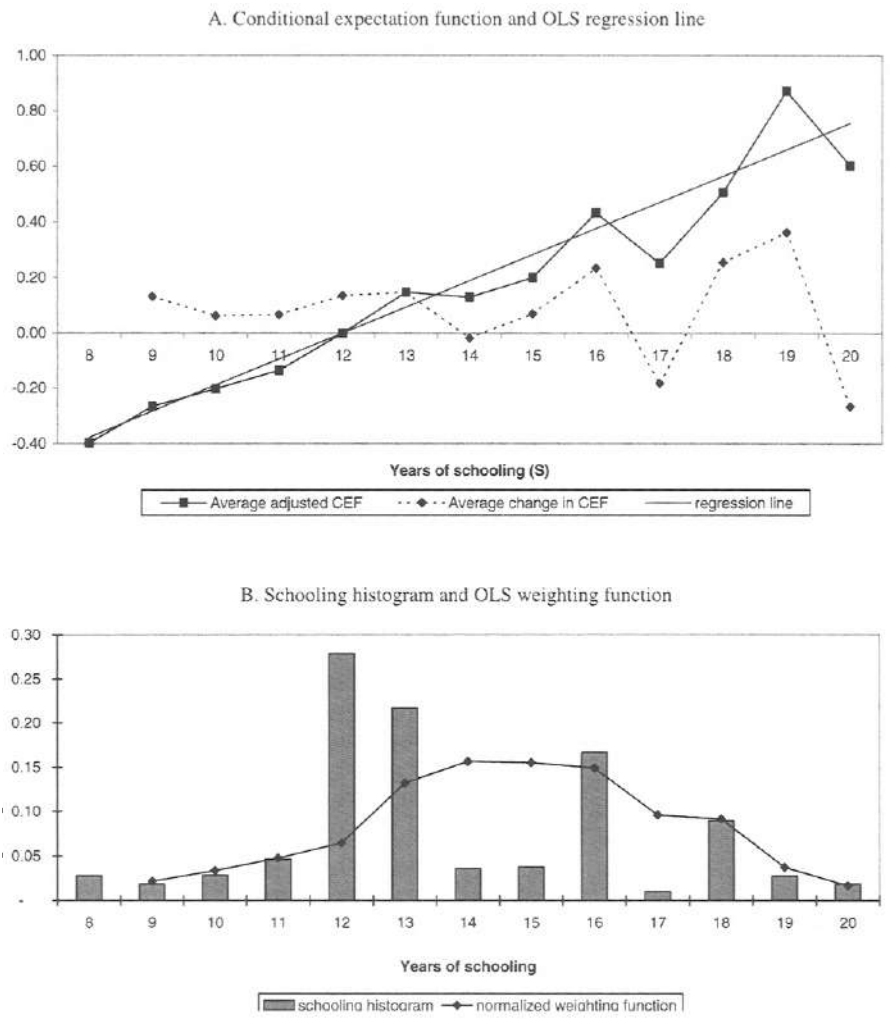

Fig. 3. (A) The conditional expectation function (CEF) of log weekly earnings given schooling, adjusted for covariates as described in the text. Also plotted is the average change in the CEF and the OLS regression line. (B) The schooling histogram and OLS weighting function. Data are for men aged 40–49 in the 1990 Census.

$P[S \geq S \mid X_i]$ equals 0 or 1. This includes values of X_i where S_i does not vary across observations, because $P[S \geq S \mid X_i] = 1$ if $P[S_i = S \mid X_i] = 1$.

The weighting function is illustrated in Fig. 3 using data from the 1990 Census. The top panel plots an estimate of the earnings-schooling CEF, i.e., average log weekly wages against years of schooling for men with 8–20 years of schooling, adjusted for covariates. In other words, the plot shows $E\{E[Y_i \mid X_i, S_i = S]\}$, plotted against S. Years of schooling

are not recorded in the 1990 Census and were therefore imputed from categorical schooling variables as described in the appendix. The X-variables are race (white, non-white), age (40–49), and state of birth. The covariates in this case are similar to those used in some of the specifications in the Angrist and Krueger (1991) study of the returns to schooling, although the data underlying this figure are more recent.

The dotted line in the figure plots the change in $E\{E[Y_i \mid X_i, S_i = S]\}$ with S. This is the covariate-adjusted difference in average log weekly wages at each schooling increment,

$$\rho_S \equiv E\{E[Y_i \mid X_i, S_i = S] - E[Y_i \mid X_i, S_i = S - 1]\} = \sum_X \rho_{SX} P(X_i = X).$$

For example, the first point on the dotted line is an estimate of $\rho_9 - \rho_8$, which is the average difference in earnings between those with 9 years of schooling and those with 8 years of schooling, adjusting for differences in the distribution of X_i between the two schooling groups.[28] The returns measured in this way are remarkably stable until 13 years of schooling, but quite variable after that and sometimes even negative.

The straight line in the figure is the OLS regression line obtained from fitting Eq. (1) with a saturated model for X_i (in other words, the model includes a full set of dummies d_{iX}, which equal one when $X_i = X$ for every value X; the OLS estimate of ρ in this case is 0.094). This parameterization satisfies assumption (30), i.e., $E[S_i \mid X_i]$ is linear. The figure illustrates the sense in which OLS captures the average return. The OLS weighting function for each value of S_i is plotted in the lower panel, along with the histogram of schooling.[29] Like the distribution of schooling itself, the OLS weighting scheme puts the most weight on values between 12 and 16. It is interesting to note, however, that while the histogram of schooling is bimodal, the weighting function is smoother and unimodal. Moreover, the population average of ρ_S, i.e., the weighted average of the covariate-adjusted return using the schooling histogram, $\sum_S \rho_S P(S_i = S)$, is 0.144, which is considerably larger than the OLS estimate. This is because about half of the sample has 12–13 years of schooling, where the returns are 0.136 and 0.148. The OLS weighting function gives more weight than the histogram to other schooling values, like 14, 15, and 17, where the returns are small and even negative.

2.3.2. Matching instead of regression

The previous section shows how regression produces a weighted average of covariate-specific effects for each value of the causing variable. The empirical consequences of the OLS weighting scheme in any particular application depend on the distribution of regressors and the amount of heterogeneity in the causal effect of interest. Matching methods provide an alternative estimation strategy that affords more control over the weighting scheme used to produce average causal effects. Matching methods also have the advantage

[28] The unadjusted difference in average wages is $\{E[Y_i \mid S_i = S] - E[Y_i \mid S_i = S - 1]\}$, which equals $E[E(Y_i \mid X_i, S_i = S) \mid S_i = S] - E[E(Y_i \mid X_i, S_i = S - 1) \mid S_i = S - 1]$.

[29] Since the regression model has covariates, the weights vary with X_i as well as for each schooling increment. The average weighting function plotted in the figure is $\sum_X \mu_{SX} P(X_i = X)$.

of making the comparisons that are used for statistical identification transparent. Matching is most practical in cases where the causing variable takes on two values, as in the union status and military service examples discussed previously.

Again, we use the example of estimating the effect of military service to illustrate this technique. Angrist (1998) reported matching and regression estimates of the effects of voluntary military service on civilian earnings. As in the Vietnam study, the potential outcomes are Y_{i0}, denoting what someone would earn if they did not serve in the military, and Y_{1i} denoting earnings as a veteran. Since $Y_{1i} - Y_{0i}$ is not constant, and we never observe both potential outcomes for any one person, it makes sense to focus on average effects. One possibility is the "average treatment effect," $E[Y_{1i} - Y_{0i}]$, but this is not usually the first choice in studies of this kind since people who serve in the military tend to have personal characteristics that differ, on average, from those of people who did not serve. The manpower policy innovations that are typically contemplated affect those individuals who either now serve or who might be expected to serve in the future. For example, between 1989 and 1992, the size of the military declined sharply because of increasing enlistment standards. Policy makers would like to know whether the people who would have served under the old rules but are unable to enlist under the new rules were hurt by the lost opportunity for service. This sort of reasoning leads researchers to try to estimate the "effect of treatment on the treated," which is $E[Y_{1i} - Y_{0i} \mid D_i = 1]$ in our notation.[30]

As in the study of Vietnam veterans, simply comparing the earnings of veterans and non-veterans is unlikely to provide a good estimate of the effect of military service on veterans. The comparison by veteran status is

$$E[Y_{1i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0]$$

$$= E[Y_{1i} - Y_{0i} \mid D_i = 1] + \{E[Y_{0i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0]\}.$$

This is the average causal effect of military service on veterans, $E[Y_1 - Y_0 \mid D = 1]$, plus a bias term attributable to the fact that the earnings of non-veterans are not necessarily representative of what veterans would have earned had they not served in the military. For example, veterans may have higher earnings simply because they must have higher test scores and be high school graduates to meet military screening rules.

The bias term in naive comparisons goes away if D_i is randomly assigned because then D_i will then be independent of Y_{0i} and Y_{1i}. Since voluntary military service is not randomly assigned (and there is no longer a draft lottery), Angrist (1998) used matching and regression techniques to control for observed differences between veterans and non-veterans who applied to get into the all-volunteer forces between 1979 and 1982. The motivation for a control strategy in this case is the fact that the military screens applicants to the armed forces primarily on the basis of age, schooling, and test scores, characteristics that are

[30] Heckman and Robb (1985) discuss the rationale for estimating effects on the treated when evaluating subsidized training programs.

observed in the Angrist (1998) data. Identification in this case is based on the claim that after conditioning on all of the observed characteristics that are known to affect veteran status, veterans and non-veterans are comparable in the sense that

$$\mathrm{E}[Y_{0i} \mid X_i, D_i = 1] = \mathrm{E}[Y_{0i} \mid X_i, D_i = 0]. \tag{35}$$

This assumption seems plausible for two reasons. First, the non-veterans who provide observations on Y_{0i} did in fact apply to get into the military. Second, selection for military service from the pool of applicants is based almost entirely on variables that are observed and included in the X-variables. Variation in veteran status conditional on X_i comes solely from the fact that some qualified applicants nevertheless fail to enlist at the last minute. Of course, the considerations that lead a qualified applicant to "drop out" of the enlistment process could be related to earnings potential, so assumption (35) is clearly not guaranteed.

Given assumption (35), the effect of treatment on the treated can be constructed as follows:

$$\begin{aligned}\mathrm{E}[Y_{1i} - Y_{0i} \mid D_i = 1] &= \mathrm{E}\{\mathrm{E}[Y_{1i} \mid X_i, D_i = 1] - \mathrm{E}[Y_{0i} \mid X_i, D_i = 1] \mid D_i = 1\} \\ &= \mathrm{E}\{\mathrm{E}[Y_{1i} \mid X_i, D_i = 1] - \mathrm{E}[Y_{0i} \mid X_i, D_i = 0] \mid D_i = 1\} = \mathrm{E}[\delta_X \mid D_i = 1], \end{aligned} \tag{36}$$

where

$$\delta_X = \mathrm{E}[Y_i \mid X_i, D_i = 1] - \mathrm{E}[Y_i \mid X_i, D_i = 0].$$

Here δ_X is a random variable that represents the set of differences in mean earnings by veteran status corresponding to each value taken on by X_i. This is analogous to the random coefficient ρ_X that was defined for the schooling problem. Note, however, that since D_i is binary, the response function in this case is automatically linear in D_i.

The matching estimator in Angrist (1998) uses the fact that X_i is discrete to construct (36), which can also be written

$$\mathrm{E}[Y_{1i} - Y_{0i} \mid D_i = 1] = \sum_X \delta_X \mathrm{P}(X_i = X \mid D_i = 1), \tag{37}$$

where $\mathrm{P}(X_i = X \mid D = 1)$ is the probability mass function for X_i given $D_i = 1$ and the summation is over the values of X_i.[31] In this case, X_i takes on values determined by all possible combinations of year of birth, AFQT test-score group,[32] year of application to the military, and educational attainment at the time of application.

Naive comparisons clearly overestimate the benefit of military service. This can be seen in Table 6, which reports differences-in-means, matching, and regression estimates of the effect of voluntary military service on the 1988–1991 Social Security-taxable earnings of men who applied to join the military between 1979 and 1982. The matching estimates were constructed from the sample analog of (37), i.e., from covariate-value-specific differ-

[31] This matching estimator is discussed by Rubin (1977) and used by Card and Sullivan (1988) to estimate the effect of subsidized training on employment.

[32] This is the Armed Forces Qualification Test, used by the military to screen applicants.

Table 6
Matching and regression estimates of the effects of voluntary military service[a]

Race	Average earnings in 1988–1991 (1)	Differences in means by veteran status (2)	Matching estimates (3)	Regression estimates (4)	Regression minus matching (5)
Whites	14537	1233.4	−197.2	−88.8	108.4
		(60.3)	(70.5)	(62.5)	(28.5)
Non-whites	11664	2449.1	839.7	1074.4	234.7
		(47.4)	(62.7)	(50.7)	(32.5)

[a] Notes: Adapted from Angrist (1998, Tables II and V). Standard errors are reported in parentheses. The tables shows estimates of the effect of voluntary military service on the 1988–1991 Social Security-taxable earnings of men who applied to enter the armed forces between 1979 and 1982. The matching and regression estimates control for applicants' year of birth, education at the time of application, and AFQT score. There are 128,968 whites and 175,262 non-whites in the sample.

ences in earnings, δ_X, weighted to form a single estimate using the distribution of covariates among veterans. Although white veterans earn \$1233 more than non-veterans, this difference becomes negative once the adjustment for differences in covariates is made. Similarly, while non-white veterans earn \$2449 more than non-veterans, controlling for covariates reduces this to \$840.

Table 6 also reports regression estimates of the effect of voluntary service, controlling for exactly the same covariates used in the matching estimates. These are estimates of δ_r in the equation

$$Y_i = \sum_X d_{iX}\beta_X + \delta_r D_i + e_i, \tag{38}$$

where β_X is a regression-effect for $X_i = X$ and δ_r is the regression treatment effect. This corresponds to a saturated model for X_i. Despite the fact that the matching and regression estimates control for the same variables, the regression estimates are significantly larger than the matching estimates for both whites and non-whites.[33] The reason the regression estimates are larger than the matching estimates is that the two estimation strategies use different weighting schemes. While the matching estimator combines covariate-value-specific estimates, δ_X, to produces an estimate of the effect of treatment on the treated, regression produces a variance-weighted average of these effects. To see this, note that since D_i is binary and $\mathrm{E}[D_i \mid X_i]$ is linear, formula (33) from the previous section implies

$$\delta_r = \mathrm{E}[(D_i - \mathrm{E}[D_i \mid X_i])^2 \delta_X]/\mathrm{E}[(D_i - \mathrm{E}[D_i \mid X_i])^2] = \mathrm{E}[\sigma_D^2(X_i)\delta_X]/\mathrm{E}[\sigma_D^2(X_i)],$$

But in this case, $\sigma_D{}^2(X_i) = \mathrm{P}(D_i = 1 \mid X_i)(1 - \mathrm{P}(D_i = 1 \mid X_i))$, so

[33] The formula for the covariance of regression and matching estimates is derived in Angrist (1998, p. 274).

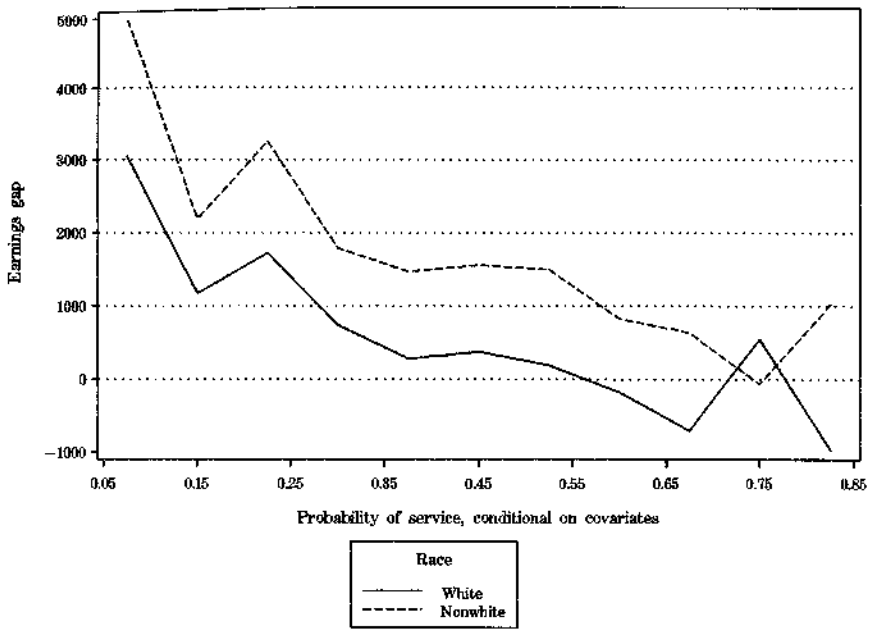

Fig. 4. Effects of voluntary military service on earnings in 1988–1991, plotted by race and probability of service, conditional on covariates. The earnings data are from Social Security administrative records.

$$\delta_r = \frac{\sum_X \delta_X[\mathrm{P}(D_i = 1 \mid X_i = X)(1 - \mathrm{P}(D_i = 1 \mid X_i = X))]\mathrm{P}(X_i = X)}{\sum_X [\mathrm{P}(D_i = 1 \mid X_i = X)(1 - \mathrm{P}(D_i = 1 \mid X_i = X))]\mathrm{P}(X_i = X)}.$$

In other words, regression weights each covariate-specific treatment effect by $\mathrm{P}(X_i = X \mid D_i = 1)(1 - \mathrm{P}(X_i = X \mid D_i = 1))$. In contrast, the matching estimator, (37), can be written

$$\mathrm{E}[Y_{1i} - Y_{0i} \mid D_i = 1] = \frac{\sum_X \delta_X \mathrm{P}(D_i = 1 \mid X_i = X)\mathrm{P}(X_i = X)}{\sum_X \mathrm{P}(D_i = 1 \mid X_i = X)\mathrm{P}((X_i = X)}.$$

because $\mathrm{P}(X_i = X \mid D_i = 1) = \mathrm{P}(D_i = 1 \mid X_i = X)\mathrm{P}(X_i = X)/\mathrm{P}(D_i)$.

The weights underlying $\mathrm{E}[Y_{1i} - Y_{0i} \mid D_i = 1]$ are proportional to the probability of veteran status at each value of the covariates. So the men most likely to serve get the most weight in estimates of the effect of treatment on the treated. In contrast, regression estimation weights each of the underlying treatment effects by the conditional variance of treatment status, which in this case is maximized when $\mathrm{P}(D_i = 1 \mid X_i = X) = 1/2$. Of course, the difference in weighting schemes is of no importance if the effect of interest

does not vary with X_i. But Fig. 4, which plots X-specific estimates (δ_X) of the effect of veteran status on average 1988–1991 earnings against $P[D_i = 1 \mid X_i = X]$, shows that the men who were most likely to serve in the military benefit least from their service. This fact leads matching estimates of the effect of military service to be smaller than regression estimates based on the same vector of controls.

2.3.3. Matching using the propensity score

It is easy to construct a matching estimator based on (37) when, as in Angrist (1998), the conditioning variables are discrete and the sample has many observations at almost every value taken on by the vector of explanatory variables. What about situations where X_i is continuous, so that exact matching is not practical? Problems involving more finely distributed X-variables are often solved by aggregating values to make coarser groupings or by pairing observations that have similar, though not necessarily identical, values. See Cochran (1965), Rubin (1973), or Rosenbaum (1995, Chapter 3) for discussions of this approach. More recently, Deaton and Paxson (1998) used non-parametric methods to accommodate continuous-valued control variables in a matching estimator.

The problem of how to aggregate the X-variables also motivates a matching method first developed in a series of papers by Rosenbaum and Rubin (1983, 1984, 1985). These papers show that full control for covariates can be obtained by controlling solely for a function of X_i called the propensity score, which is simply the conditional probability of treatment, $p(X_i) \equiv P(D_i = 1 \mid X_i)$. The formal result underlying this approach says that if conditioning on X_i eliminates selection bias,

$$E[Y_{0i} \mid X_i, D_i = 1] = E[Y_{0i} \mid X_i, D_i = 0],$$

then it must also be true that conditioning on $p(X_i)$ eliminates selection bias:

$$E[Y_{0i} \mid p(X_i,), D_i = 1] = E[Y_{0i} \mid e(X_i), D_i = 0].$$

This leads to the following modification of (36):

$$E[Y_{1i} - Y_{0i} \mid D_i = 1] = E\{E[Y_{1i} \mid X_i, D_i = 1] - E[Y_{0i} \mid X_i, D_i = 1] \mid D_i = 1\}$$

$$= E\{E[Y_{1i} \mid p(X_i), D_i = 1] - E[Y_{0i} \mid p(X_i), D_i = 0] \mid D_i = 1\}.$$

Of course, to make this expression into an estimator, the propensity score $p(X_i)$ must first be estimated. The practical value of this result is that in some cases, it may be easier to estimate $p(X_i)$ and then condition on the estimates of $p(X_i)$ than to condition on X_i directly. For example, even if X_i is continuous, $p(X_i)$ may have some "flat spots", or we may have some prior information about $p(X_i)$. The propensity score approach is also conceptually appealing because it focuses attention on variables that are related to the regressor of interest. Although Y_i may vary with X_i in complicated ways, this is only of concern for values of X_i where $p(X_i)$ varies as well.

An example using the propensity score in labor economics is Dehejia and Wahba's (1995) reanalysis of the National Supported Work (NSW) training program studied by

Lalonde (1986). The NSW provided training to different groups of "hard-to-employ" men and women in a randomized demonstration project. Lalonde's study uses observational control groups from the Current Population Survey (CPS) and the Panel Study of Income Dynamics (PSID) to look at whether econometric methods are likely to generate conclusions similar to those found in the experimental study. One hurdle facing the non-experimental investigator attempting to construct a control group for trainees is how to control for lagged earnings. As we noted earlier, controlling for lagged earnings is important since participants in government training programs are often observed to experience a decline in earnings before entering the program (see, e.g., Ashenfelter and Card, 1985, and the chapter on training by Heckman, Lalonde, and Smith in this volume).

Lalonde (1986) found that non-experimental methods based on regression models, including models with fixed effects and control for lagged earnings, fail to replicate the NSW experimental findings. Using the same observational control groups as Lalonde (1986), Dehejia and Wahba (1995) control for lagged earnings and other covariates by first estimating a logit model that relates participation in the program to the covariates and two lags of earnings. Following an example by Rosenbaum and Rubin (1984), they then divide the sample into quintiles on the basis of fitted values from this logit, i.e., based on estimates of the propensity score. The overall estimate of the effect of treatment on the treated is the difference between average trainee and average control earnings in each quintile, weighted by the number of trainees in the quintile and summed across quintiles. The estimates produced using this method are similar to those based on the experimental random assignment (and apparently more reliable than regression estimates). It should be clear, however, that use of propensity score methods requires a number of decisions about how to model and control for the score. There is little in the way of formal statistical theory to guide this process, and the question of whether propensity score methods are better than other methods remains open. See Heckman et al. (1997) for further empirical evidence, and Hahn (1998) for recent theoretical results on efficiency considerations in these models.

2.3.4. Interpreting instrumental variables estimates

The discussion of IV in Section 2.2.3 used the example of veteran status, with two potential outcomes and a constant causal effect, $Y_{1i} - Y_{0i} = \delta$. What is the interpretation of an IV estimate when the constant-effects assumption is relaxed? We begin with a model where the causing variable is binary, as in the veteran status example, turning afterwards to a more general model. As before, the discussion is initially limited to the Wald estimator since this is an important and easily-analyzed IV estimator.

Without the constant-effects assumption, we can write the observed outcome, Y_i, in terms of potential outcomes as

$$Y_i = Y_{i0} + (Y_{1i} - Y_{0i})D_i = \beta_0 + \delta_i D_i + \eta_i, \tag{39}$$

where $\beta_0 \equiv \mathrm{E}[Y_{i0}]$ and $\delta_i \equiv Y_{1i} - Y_{0i}$ is the heterogeneous causal effect. The expression after the second equals sign is a "random-coefficients" version of the causal model in Section 2.3.3 (see Eq. (23)). To facilitate the discussion of IV, we also introduce some

notation for the first-stage relationship between the causing variable, D_i, and the binary instrument, Z_i. To allow for as much heterogeneity as possible, the first stage equation is written in a manner similar to (39):

$$D_i = D_{i0} + (D_{1i} - D_{0i})Z_i = \pi_0 + \pi_{1i}Z_i + \nu_i, \tag{40}$$

where $\pi_0 \equiv \mathrm{E}[D_{i0}]$ and $\pi_{1i} \equiv (D_{1i} - D_{0i})$ is the causal effect of the *instrument* on D_i. In the draft lottery example, D_{0i} tells us whether i would serve in the military if not draft-eligible and D_{1i} tells us whether i would serve when draft-eligible. The effect of draft-eligibility on D_i is the difference between these two potential treatment assignments.

The principle identifying assumption in this setup is that the vector of potential outcomes and potential treatment assignments is jointly independent of the instrument. Formally,

$$\{Y_{1i}, Y_{0i}, D_{1i}, D_{0i}\} \amalg Z_i,$$

where $\amalg$ is notation for statistical independence (see, e.g., Dawid, 1979, or Rosenbaum and Rubin, 1983).[34] In the lottery example, Z_i is clearly independent of $\{D_{0i}, D_{1i}\}$ since Z_i was randomly assigned. As noted in Section 2.2.3, however, independence of $\{Y_{0i}, Y_{1i}\}$ and Z_i is not guaranteed by randomization since Y_{0i} and Y_{1i} refer to potential outcomes under alternative assignments of *veteran status* and not Z_i itself. Even though Z_i was randomly assigned, so the relationship between Z_i and Y_i is causal, in principle there might be reasons other than veteran status for an effect of draft-eligibility on earnings. The independence assumption, which is similar to the assumption that Z_i and η_i are uncorrelated in the constant-effects model, rules this possibility out.

A second assumption that is useful here, and one that does not arise in a constant-effects setting, is that either $\pi_{1i} \geq 0$ for all i or $\pi_{1i} \leq 0$ for all i. This *monotonicity* assumption, introduced by Imbens and Angrist (1994), means that while the instrument may have no effect on some people, it must be the case that the instrument acts in only one direction, either $D_{1i} \geq D_{0i}$ or $D_{1i} \leq D_{0i}$ for all i. In what follows, we assume $D_{1i} \geq D_{0i}$ for all i. In the draft-lottery example, this means that although draft-eligibility may have had no effect on the probability of military service for some men, there is no one who was actually kept out of the military by being draft-eligible. Without monotonicity, instrumental variables estimators are not guaranteed to estimate a weighted average of the underlying causal effects, $Y_{1i} - Y_{0i}$.

Given independence and monotonicity, the Wald estimator in this example can be interpreted as the effect of veteran status on those whose treatment status was changed by the instrument. This parameter is called the local average treatment effect (LATE; Imbens and Angrist, 1994), and can be written as follows:

$$\frac{\mathrm{E}[Y_i \mid Z_i = 1] - \mathrm{E}[Y_i \mid Z_i = 0]}{\mathrm{E}[D_i \mid Z_i = 1] - \mathrm{E}[D_i \mid Z_i = 0]} = \mathrm{E}[Y_{1i} - Y_{0i} \mid D_{1i} > D_{0i}] = \mathrm{E}[\delta_i \mid \pi_{1i} > 0].$$

[34] The independence assumption using random-coefficients notation is $\{\delta_i, \eta_i, \pi_{1i}, \nu_i\} \amalg Z_i$.

Thus, IV estimates of effects of military service using the draft lottery estimate the effect of military service on men who served because they were draft-eligible, but would not otherwise have served.[35] This obviously excludes volunteers and men who were exempted from military service for medical reasons, but it includes men for whom the draft policy was binding. Much of the debate over compulsory military service focused on draftees, so LATE is clearly a parameter of policy interest in the Vietnam context.

The LATE parameter can be linked to the parameters in traditional econometric models for causal effects. One commonly used specification for dummy endogenous regressors like veteran status is a latent-index model (see, e.g., Heckman, 1978), where

$$D_i = 1 \quad \text{if } \gamma_0 + \gamma_1 Z_i > \nu_i \quad \text{and 0 otherwise,}$$

and ν_i is a random factor assumed to be independent of Z_i. This specification can be motivated by comparisons of utilities and costs under alternative choices. In the notation of Eq. (40), the latent-index model characterizes potential treatment assignments as

$$D_{0i} = 1 \text{ if } [\gamma_0 > \nu_i] \quad \text{and} \quad D_{1i} = 1 \text{ if } [\gamma_0 + \gamma_1 > \nu_i].$$

Note that in this model, monotonicity is automatically satisfied since γ_1 is a constant. Assuming $\gamma_1 > 0$,

$$\mathrm{E}[Y_{1i} - Y_{0i} \mid D_1 > D_{0i}] = \mathrm{E}[Y_{1i} - Y_{0i} \mid \gamma_0 + \gamma_1 > \nu_i > \gamma_0],$$

which is a function of the structural first-stage parameters, γ_0 and γ_1. The LATE parameter is representative of a larger group the larger is the first-stage parameter, γ_1.

LATE can also be compared with the effect of treatment on the treated for this problem, which depends on the same first-stage parameters and the marginal distribution of Z_i. Note that in the latent-index specification, $D_i = 1$ in one of two ways: either $\gamma_0 > \nu_i$, in which case the instrument does not matter, or $\gamma_0 + \gamma_1 > \nu_i > \gamma_0$ and $Z_i = 1$. Since these two possibilities partition the group with $D_i = 1$, we can write

$$\begin{aligned}
\mathrm{E}[Y_{1i} - Y_{0i} \mid D_i = 1] &= \mathrm{P}(D_i = 1)^{-1} \\
&\times \{\mathrm{E}[Y_{1i} - Y_{0i} \mid \gamma_0 + \gamma_1 > \nu_i > \gamma_0, Z_i = 1]\mathrm{P}(\gamma_0 + \gamma_1 > \nu_i > \gamma_0, Z_i = 1) \\
&+ \mathrm{E}[Y_{1i} - Y_{0i} \mid \gamma_0 > \nu_i]\mathrm{P}(\gamma_0 > \nu_i)\} \\
&= \mathrm{P}(D_i = 1)^{-1} \times \{\mathrm{E}[Y_{1i} - Y_{0i} \mid \gamma_0 + \gamma_1 > \nu_i > \gamma_0]\mathrm{P}(\gamma_0 + \gamma_1 > \nu_i > \gamma_0)\mathrm{P}(Z_i = 1) \\
&+ \mathrm{E}[Y_{1i} - Y_{0i} \mid \gamma_0 > \nu_i]\mathrm{P}(\gamma_0 > \nu_i)\}.
\end{aligned}$$

[35] Proof of the LATE result: $\mathrm{E}[Y_i \mid Z_i = 1] = \mathrm{E}[Y_{i0} + (Y_{1i} - Y_{0i})D_i \mid Z_i = 1]$, which equals $\mathrm{E}[Y_{i0} + (Y_{1i} - Y_{0i})D_{1i}]$ by independence. Likewise $\mathrm{E}[Y_i \mid Z_i = 0] = \mathrm{E}[Y_{i0} + (Y_{1i} - Y_{0i})D_{0i}]$, so the numerator of the Wald estimator is $\mathrm{E}[(Y_{1i} - Y_{0i})(D_{1i} - D_{0i})]$. Monotonicity means $D_{1i} - D_{0i}$ equals one or zero, so $\mathrm{E}[(Y_{1i} - Y_{0i})(D_{1i} - D_{0i})] = \mathrm{E}[Y_{1i} - Y_{0i} \mid D_{1i} > D_{0i}]\mathrm{P}[D_{1i} > D_{0i}]$. A similar argument shows $\mathrm{E}[D_i \mid Z_i = 1] - \mathrm{E}[D_i \mid Z_i = 0] = \mathrm{E}[D_{1i} - D_{0i}] = \mathrm{P}[D_{1i} > D_{0i}]$.

This shows that the effect on the treated is a weighted average of LATE and the effect on men whose treatment status is unaffected by the instrument.[36] Note, however, that although LATE equals the Wald estimator, the effect on the treated is not identified in this case without additional assumptions (see, e.g., Angrist and Imbens, 1991).

Interpreting IV estimates with cardinal variables. So far the discussion of IV has focused on models with a binary regressor. What does the Wald estimator estimate when the regressor takes on more than two values, like schooling? As in the discussion of regression in Section 2.2.1, suppose the causal relationship of interest is characterized by a function that describes exactly what a given individual would earn if they obtained different levels of education. This relationship is person-specific, so we write $f_i(S)$ to denote the earnings or wage that i would receive after obtaining S years of education. The observed earnings level is $Y_i = f_i(S_i)$.

Again, it is useful to have a general notation for the first-stage relationship between S_i and Z_i:

$$S_i = S_{0i} + (S_{1i} - S_{0i})Z_i = \kappa_0 + \kappa_{1i}Z_i + \nu_i, \tag{41}$$

where S_{0i} is the schooling i would get if $Z_i = 0$, S_{1i} is the schooling i would get if $Z_i = 1$, and $\kappa_0 \equiv \mathrm{E}[S_{0i}]$. In random-coefficients notation, the causal effect of Z_i on S_i is $\kappa_{1i} \equiv S_{1i} - S_{0i}$. To make this concrete, suppose the instrument is a dummy for being born in the second, third, or fourth quarter of the year, as for the Wald estimate in Angrist and Krueger (1991, Table 3). Since compulsory attendance laws allow people to drop out of school on their birthday (typically the 16th) and most children enter school in September of the year they turn 6, pupils born later in the year are kept in school longer than those born earlier. In this example, S_{0i} is the schooling i would get if born in the first quarter and S_{1i} is the schooling i would get if born in a later quarter.

Now the independence assumption is $\{f_i(S), S_{1i}, S_{0i}\} \perp\!\!\!\perp Z$, and the monotonicity assumption is $S_{1i} \geq S_{0i}$. This means the instrument is independent of what an individual *could* earn with schooling level S, and independent of the random elements in the first stage.[37] Using the independence assumption and Eq. (41) to substitute for S_i, the Wald estimator can be written

$$\frac{\mathrm{E}[f_i(S_i) \mid Z_i = 1] - \mathrm{E}[f_i(S_i) \mid Z_i = 0]}{\mathrm{E}[S_i \mid Z_i = 1] - \mathrm{E}[S_i \mid Z_i = 0]} = \frac{\mathrm{E}[f_i(S_{1i}) - f_i(S_{0i})]}{\mathrm{E}[S_{1i} - S_{0i}]}$$

$$= \mathrm{E}\{\omega_i[(f_i(S_{1i}) - f_i(S_{0i}))/(S_{1i} - S_{0i})]\}, \tag{42}$$

where $\omega_i \equiv (S_{1i} - S_{0i})/\mathrm{E}[S_{1i} - S_{0i}]$. This is a weighted average arc-slope of $f_i(S)$ on the interval $[S_{0i}, S_{1i}]$. We can simplify further using the fact that $f_i(S_{1i}) =$

[36] Note that $\mathrm{P}[\gamma_0 + \gamma_1 > \nu_i > \gamma_0]\mathrm{P}[Z_i = 1] + \mathrm{P}[\gamma_0 > \nu_i] = (\mathrm{E}[D_i \mid Z_i = 1] - \mathrm{E}[D_i \mid Z_i = 0])\mathrm{P}(Z_i = 1) + \mathrm{E}[D_i \mid Z_i = 0] = \mathrm{P}[D_i = 1]$, so the weights sum to one. In the special case where $\mathrm{P}[\gamma_0 > \nu_i] = 0$ for everyone, LATE and the effect of treatment on the treated are the same.

[37] For example, if $f_i(S) = \beta_0 + \rho_i S + \eta_i$, then we assume $\{\rho_i, \eta_i, \kappa_{1i}, \nu_i\}$ are independent of Z_i.

$f_i(S_{0i}) + f_i'(S_i^*)(S_{1i} - S_{0i})$, for some S_i^* in the interval $[S_{0i}, S_{1i}]$.[38] Now we can write the Wald estimator as an average derivative:

$$\frac{E[f_i(S_{1i}) - f_i(S_{0i})]}{E[S_{1i} - S_{0i}]} = \frac{E[(S_{1i} - S_{0i})f_i'(S_i^*)]}{E[S_{1i} - S_{0i}]} = E[\omega_i f_i'(S_i^*)]. \tag{43}$$

Given the monotonicity assumption, ω_i is positive for everyone, so the Wald estimator is a weighted average of individual-specific slopes at a point in the interval $[S_{0i}, S_{1i}]$. The weight each person gets is proportional to the size of the causal effect of the instrument on him or her. The range of variation in $f_i(S)$ summarized by this average is always between S_{0i} and S_{1i}.

Angrist et al. (1995) note that the Wald estimator can be characterized more precisely in a number of important special cases. First, suppose that the effect of the instrument is the same for everybody, i.e., κ_{1i} is constant. Then we obtain the average derivative $E[f_i'(S_i^*)]$, and no weighting is involved. If $f_i(S)$ is linear in S, as in Section 2.2.1, but with a random coefficient, ρ_i then the Wald estimator is a weighted average of the random coefficient: $E[(S_{1i} - S_{0i})\rho_i]/E[S_{1i} - S_{0i}]$. If κ_{1i} is constant *and* $f_i(S)$ is linear, then the Wald estimator is the population average slope, $E[\rho_i]$.

Another interesting special case is when $f_i(S)$ is a quadratic function of S, as in Lang (1993) and Card's (1995) parameterization of a structural human-capital earnings function. The quadratic function captures the notion that returns to schooling decline as schooling increases. Note that for a quadratic function, the point of linearization is always $S_i^* = (S_{1i} + S_{0i})/2$. The Wald estimator is therefore

$$E[\omega_i f_i'((S_{1i} + S_{0i})/2)],$$

i.e., a weighted average of individual slopes at the midpoint of the interval $[S_{0i}, S_{1i}]$ for each person. The fact that the weights are proportional to $S_{1i} - S_{0i}$ sometimes has economic significance. In the Card and Lang models, for example, the first-stage effect, $S_{1i} - S_{0i}$, is assumed to be proportional to individual discount rates. Since people with higher discount rates get less schooling and the schooling-earnings relationship has been assumed to be concave, this tends to make the Wald estimate higher than the population average return. Lang (1993) called this phenomenon "discount rate bias".

In some applications, it is interesting to characterize the range of variation captured by the Wald estimator further. Returning to (42), which describes the estimator as a weighted average of slopes in the interval $[S_{0i}, S_{1i}]$, it seems natural to ask which values are most likely to be covered by this interval. For example, does $[S_{0i}, S_{1i}]$ usually cover 12 years of education, or is it more likely to cover 16 years? The probability $S \in [S_{0i}, S_{1i}]$ is $P[S_{1i} \geq S \geq S_{0i}]$. Because S_i is discrete, it is easier to work with $P[S_{1i} > S \geq S_{0i}]$, since this can be expressed as

[38] Here we assume that $f_i(S)$ is continuously differentiable with domain equal to a subset of the real line.

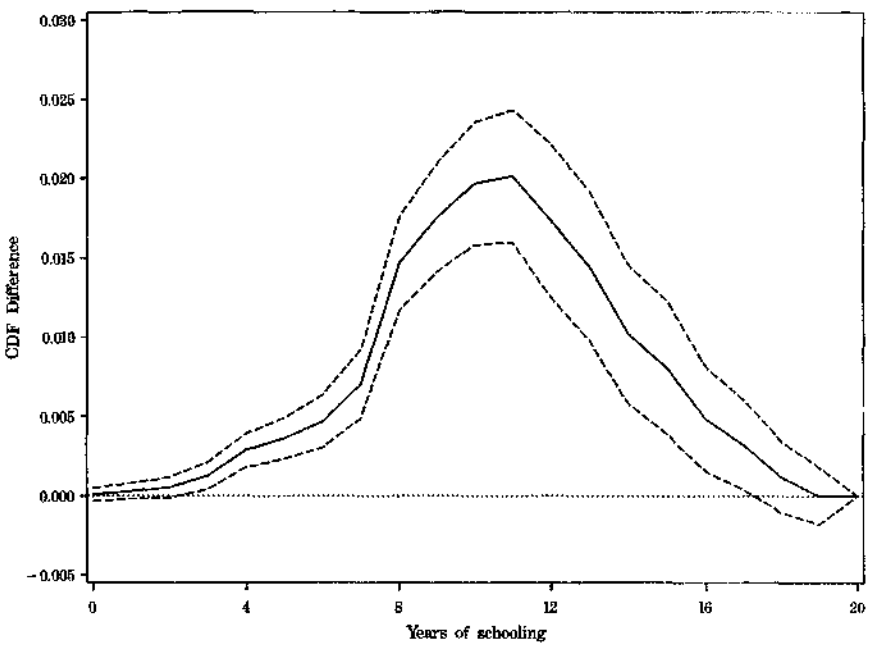

Fig. 5. First quarter–fourth quarter difference in schooling CDFs, for men born 1930–1939 in the 1980 Census. The dotted lines are 95% confidence intervals.

$$\mathrm{P}[S_{1i} > S \geq S_{0i}] = \mathrm{P}[S_{1i} > S] - \mathrm{P}[S_{0i} > S] = \mathrm{P}[S_i \leq S \mid Z_i = 0] - \mathrm{P}[S_i \leq S \mid Z_i = 1]. \tag{44}$$

This is the difference in the cumulative distribution function (CDF) of schooling with the instrument switched off and on. The schooling values where the CDF-gap is largest are those most likely to be covered by the interval $[S_{0i}, S_{1i}]$, and therefore most often represented in the Wald/weighted average.

Angrist and Imbens (1995) used Eq. (44) to interpret the Wald estimates of the returns to schooling reported by Angrist and Krueger (1991).[39] They report a Wald estimate based on first quarter/fourth quarter differences in log weekly wages and years of schooling using data on men born 1930–1939 in the 1980 Census. Their Wald estimate is 0.089, and the corresponding OLS estimate is 0.07. The first quarter/fourth quarter difference in CDFs is plotted in Fig. 5. The difference is largest in the 8–14 years-of-schooling range. This is not surprising since compulsory attendance laws mainly affect high school students, i.e., those with 8–12 years of education. The CDF gap for men with more than 12 years of schooling

[39] See Kling (1998) for a similar analysis of instrumental variables estimates using distance to college as an instrument for schooling.

may be caused by men who were compelled to complete high school but then attended college later.

Finally, we note that the discussion of IV in heterogeneous and non-linear models so far has ignored covariates. 2SLS estimates in heterogeneous-effects models with covariates can be interpreted in much the same way as regression estimates of models with covariates were interpreted in Section 2.3.1. That is, IV estimates in models with covariates can be thought of as producing a weighted average of covariate-specific Wald estimates as long as the model for covariates is saturated and $E[S_i \mid X_i, Z_i]$ is used as an instrument. In other cases it seems reasonable to assume that some sort of approximate weighted average is being generated, but we are unaware of a precise causal interpretation that fits all cases.[40]

2.4. Refutability

Causality can never be proved by associations in non-experimental data. But sometimes the lack of association between variables for a particular group, or the occurrence of an association between the "causing variable" and outcome variable for a group thought to be unaffected by the treatment, can cast doubt on, or even refute, a causal interpretation. R.A. Fisher (quoted in Cochran, 1965) argued that the case for causality is stronger when the causal model has many implications that appear to hold. For this reason, he suggested that scientific theories be made "complicated," in the sense that they yield many testable implications.

A research design is more likely to be successful at assessing causality if possibilities for checking collateral implications of causal processes are "built in." At one level, this involves estimating less restrictive models. A good example is Freeman's (1984) panel data study of union status, which looks separately at workers who join unions and leave unions. If unions truly raise wages of their members, then workers who move from non-union to union jobs should experience a raise, and workers who move from union to non-union jobs should experience a pay cut. Although a less restrictive model may yield imprecise estimates or be subject to different biases which render the results difficult to interpret (e.g., different unobserved variables may cause workers to join and exit union jobs), a causal story is strengthened if the results of estimating a less restrictive model are consistent with the story.

In addition to these considerations of robustness, a causal model will often yield testable predictions for sub-populations in which the "treatment effect" should *not* be observed, either because the sub-population is thought to be immune to the treatment or did not receive the treatment. Perhaps the best-known example of this type of analysis is Bound's (1989) study of the effect of Disability Insurance (DI) benefits on the labor force participation rates of older men. Earlier studies (e.g., Parsons, 1980) established an inverse

[40] A recent effort in this direction is Abadie (1998), who presents conditions under which 2SLS estimates can be interpreted as the best linear predictor for an underlying causal relationship. He also introduces a new IV estimator that always has this property for models with a single binary instrument.

relationship between the participation rate and the DI benefit-wage replacement ratio. But because the replacement ratio is a decreasing function of a worker's past earnings, Bound argued that this association may reflect pre-existing patterns of labor force participation rather than a causal response to DI benefits.[41]

To test the causal interpretation of earlier work, Bound performed two types of analyses. First, he estimated essentially the same econometric model of the relationship between employment and potential DI benefits that had been estimated previously, *except* he estimated the model for a sub-sample of older men who had never applied for DI. Because one would not expect DI benefits to provide a strong work disincentive for this sub-sample, there should be a much weaker relationship, or no relationship at all, if the causal interpretation of DI benefit coefficients is correct. Instead, he found that DI benefits had about the same effect in this sample as in a sample that included men who actually applied for and received DI benefits, suggesting that a causal interpretation of the effect of DI benefits was not warranted. Second, Bound examined the labor force behavior of men who applied for DI but were turned down. He reasoned that because men in this sub-sample were less severely disabled than men who received DI, the labor force participation rate of this sub-sample provided a "natural 'control' group" (p. 482) for predicting the upper bound of the labor force participation rate of DI recipients *had they been* denied DI benefits. Because half of the presumably healthier rejected DI applicants did not work even without receiving benefits, Bound concluded that most DI recipients did not work because they were disabled, not because DI benefits induced them to leave the labor force.

Notions of "refutability" also carry over to IV models. In Angrist and Krueger (1991) we were concerned that quarter of birth, which was the instrument for schooling, might have influenced educational attainment through some mechanism other than the interaction of school start age and compulsory schooling laws. To test this threat to a causal interpretation of the IV estimates, we examined whether quarter of birth influenced schooling or earnings for college graduates, who presumably were unaffected by compulsory schooling laws. Although quarter of birth had an effect on these outcomes for college graduates, the effect was weak and had a different pattern than that found for the less-than-college group, suggesting that compulsory schooling was responsible for the effects of quarter of birth in the less-than-college sample.

Tests of refutability may have flaws. It is possible, for example, that a subpopulation that is believed to be unaffected by the intervention is indirectly affected by it. For example, Parsons (1991) argues that rejected DI applicants are a misleading control group because they may exit the labor force to strengthen a possible appeal of their rejected application or a future re-application for DI benefits.[42] Likewise, some students who complete high school because of compulsory schooling may be induced to go on to college as a result, invalidating our 1991 test of refutability. An understanding of the institutions underlying the intervention being evaluated is necessary to assess tests of

[41] Welch (1977) provides a closely related criticism of work on Unemployment Insurance benefits.

[42] Bound (1989) considered and rejected these threats to his control group. See also Bound's (1991) response to Parsons (1991).

refutability, as well as to identify subpopulations that are immune from the intervention according to the causal story but still subject to possible confounding effects.

Lastly, there has been much recent interest in evaluating entire research designs, as in Lalonde's (1986) landmark study comparing experimental and non-experimental research methods. Only rarely, however, have experiments been conducted that can be used to validate non-experimental research strategies. Nonetheless, non-experimental research designs can still be assessed by comparing "pre-treatment" trends for the treatment and comparison group (e.g., Ashenfelter and Card, 1985; Heckman and Hotz, 1989) or by looking for effects where there should be none (e.g., Bound, 1989). We provide another illustration of this point with some new evidence on the differences-in-differences approach used in Card's (1990) immigration study.

In the summer of 1994, tens of thousands of Cubans boarded boats destined for Miami in an attempt to emigrate to the United States in a second Mariel Boatlift that promised to be almost as large as the first one, which occurred in the summer of 1980. Wishing to avoid the political fallout that accompanied the earlier boatlift, the Clinton Administration interceded and ordered the Navy to divert the would-be immigrants to a base in Guantanamo Bay. Only a small fraction of the Cuban emigres ever reached the shores of Miami. Hence, we call this event, "The Mariel Boatlift That Did not Happen."

Had the migrants been allowed to reach the United States, there is little doubt that researchers would have used this "natural experiment" to extend Card's (1990) influential study of the earlier influx of Cuban immigrants. Nonetheless, we can use this "non-event" to explore Card's research design. In particular, we can ask whether Miami's and the comparison cities' experiences were in fact similar absent the large wave of immigrants to Miami. Fig. 1, which we referred to earlier in the discussion of Card's paper, shows that non-agricultural employment growth in Miami tracks that of the four comparison cities rather well in the year before and few years after the summer of 1994. (A vertical bar indicates the date of the thwarted boatlift.) To provide a more detailed analysis by ethnic group, we followed Card and calculated unemployment rates for Whites, Blacks and Hispanics in Miami and the four comparison cities using data from the CPS Outgoing Rotation Groups. These results are reported in Table 7.

The Miami unemployment data are imprecise and variable, but still indicate a large increase in unemployment in 1994, the year the potential immigrants were diverted to Guantanamo Bay. On the other hand, 1994 was the first year the CPS redesign was implemented (see Section 3.1). We therefore take 1993 as the "pre" period and 1995 as the "post" period for a difference-in-differences comparison. For Whites and Hispanics, the unemployment rate fell in Miami and fell even more in the comparison cities between the pre and post periods, though the difference between these two changes is not significant. This is consistent with a causal interpretation of Card's (1990) results, which attributes the difference-in-differences to the effect of immigration. For blacks, however, the unemployment rate rose by 3.6 percentage points in Miami between 1993 and 1995, while it fell by 2.7 points in the comparison cities. The 6.3 point difference-in-differences estimate is on the margin of statistical significance ($t = 1.70$), and would have made it

Table 7
Unemployment rates of individuals age 16–61 in Miami and four comparison cities, 1988–1996[a]

	1988	1989	1990	1991	1992	1993	1994	1995	1996
Miami									
Whites	2.8	3.6	3.3	5.7	4.2	4.9	6.2	3.9	4.4
	(0.8)	(0.9)	(0.9)	(1.2)	(1.1)	(1.3)	(1.4)	(1.4)	(1.2)
Blacks	10.0	11.8	11.9	8.8	10.1	10.1	15.1	13.7	11.1
	(1.7)	(1.8)	(1.9)	(1.9)	(2.0)	(2.1)	(2.4)	(2.8)	(2.4)
Hispanics	5.5	7.6	7.2	9.1	10.3	8.5	9.4	8.4	8.9
	(1.4)	(1.5)	(1.4)	(1.6)	(1.7)	(1.6)	(1.8)	(1.8)	(1.6)
Comparison cities									
Whites	4.2	3.5	3.8	4.9	5.1	5.4	5.0	4.1	4.1
	(0.3)	(0.2)	(0.2)	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)
Blacks	11.3	8.4	9.6	9.6	13.6	11.5	10.9	8.8	9.3
	(0.9)	(0.8)	(0.8)	(0.9)	(1.0)	(0.9)	(0.9)	(0.8)	(0.8)
Hispanics	7.2	7.5	5.8	9.1	10.9	11.3	11.0	10.0	9.4
	(0.7)	(0.6)	(0.4)	(0.5)	(0.6)	(0.6)	(0.6)	(0.7)	(0.6)

[a] Note: Standard errors are in parentheses. The four comparison cities (Atlanta, Houston, Los Angeles, and Tampa-St. Petersburg), are the same comparison cities used by Card (1990). The reported unemployment rates are from the authors' tabulations of CPS Outgoing Rotation Groups.

look like the immigrant flow had a negative impact on Blacks in Miami in a DD study. Since there was no immigration shock in 1994, this illustrates that different labor market trends can generate spurious findings in research of this type.

3. Data collection strategies

Table 1 documents that labor economists use many different types of datasets. The renewed emphasis on quasi-experiments in empirical research places a premium on finding datasets for a particular population and time period containing certain key variables. Often this type of analysis requires large samples, because only part of the variation in the variables of interest is used in the estimation. Familiarity with datasets is as necessary for modern labor economics as is familiarity with economic theory or econometrics. Knowledge of the populations covered by the main surveys, the design of the surveys, the response rate, the variables collected, the size of the samples, the frequency of the surveys, and any changes in the surveys over time is essential for successfully implementing an empirical strategy and for evaluating others' empirical research. This section provides an overview of the most commonly used datasets and data collection strategies in labor economics.

Table 8
Commonly-used micro datasets in labor economics

Dataset	Sampling unit	Strengths	Weaknesses	Key variables
Current Population Survey (CPS), esp. March, May, and Outgoing Rotation Groups (ORG)	Household	Large samples; many years of data; many questions and supplements; basic labor force data every month; can link one survey month to another	Some questions change over time; earnings truncated and truncation point varies over time; about 80–85% non-response to income questions; mismatches in linked files	March: annual earnings and work experience, migration, cash and non-cash income, pension and health insurance coverage; May: earnings, multiple job holding, and premium pay; ORG: Labor force questions and earnings; available after 1979
Panel Study of Income Dynamics (PSID)	Household and household spin-offs	Long panel; many labor questions; low non-response rate after first few years; useful for intergenerational issues; oversamples poor families	Small for some purposes; wage rate not always available	Income sources and amounts, employment, family composition changes, demographic events, housing and food expenditures, wealth, housework time, and health status

National Longitudinal Surveys (NLS)	Individual	Concentrates on cohorts by specific age ranges; designed to be longitudinal data	Not all cohorts covered; alternate questions asked in different years; hours data are inconsistent	Labor force experience, education and training, local labor market variables, and cohort-specific questions
Census Data, esp. 1940, 1950, 1960, 1970. 1980, and 1990	Household	Gigantic samples; precise information on census tract	Wage data are noisy; Collects data on fewer variables than CPS; non-response potentially a problem	Place of birth, annual earnings, labor force status, weeks worked, month of birth, and housing variables
Survey of Income and Program Participation (SIPP)	Household	Large sample; emphasis on income and government programs	Short panel dimension; unwieldy dataset; long survey; non-response high; underreporting of program participation	Employment, education and training, program participation, assets and liabilities, migration, fertility, work schedules, child care, pension, property, and time spent out of work

3.1. Secondary datasets

The most commonly used secondary datasets in labor economics are the National Longitudinal Surveys (NLS), the Current Population Survey (CPS), the Panel Study of Income Dynamics (PSID), and the Decennial Censuses. Table 8 summarizes several features of the main secondary datasets used by labor economists. In this section we provide a more detailed discussion of the "big three" micro datasets in labor economics: the NLS, CPS and PSID. We also discuss historical comparability in the CPS and the census.

Perhaps because of its easy-to-use CD-ROM format and the breadth of its questionnaire, the National Longitudinal Surveys are popular in applied work. The NLS actually consists of six age-by-gender datasets: a cohort of 5020 "older men" (age 45–59 in 1966); a cohort of 5083 mature women (age 30–44 in 1967), a cohort of 5225 young men (age 14–24 in 1966); a cohort of 5159 young women (age 14–24 in 1968) in 1968); a cohort of 12,686 "youth" known as the NLSY (age 14–22 in 1979); and a cohort of 7035 children of respondents in the NLSY (age 0–20 in 1986).[43] Sampled individuals are interviewed annually. All but the older-men and young-men surveys continue today.

The CPS is an ongoing survey of more than 50,000 households that is conducted each month by the Census Bureau for the Bureau of Labor Statistics (BLS).[44] Sampled households are included in the survey for four consecutive months, out of the sample for 8 months, and then included for a final four consecutive months. Thus, the survey has a "rotation group" design, with new rotation groups joining or exiting the sample each month. The resulting data are used by the Bureau of Labor Statistics to calculate the unemployment rate and other labor force statistics. The CPS has a hierarchical household-family-person record structure which enables household-level and family-level analyses, as well as individual-level analyses. The design of the CPS has been copied by statistical agencies in several other countries and is similarly used to calculate labor force statistics.

In the US, regular and one-time supplements are included in the survey to collect information on worker displacement, contingent work, school enrollment, smoking, voting, and other important behaviors. In addition, annual income data from several sources are collected each month. A great strength of the CPS is that the survey began in the 1940s, so a long time-series of data are available; on the other hand, there have been several changes that affect the comparability of the data over time, and micro data are only available to researchers for years since 1964. In addition, because of its rotation group design, continuing households can be linked from one month to the next, or between years; however, individuals who move out of sampled households are not tracked, and it is possible that individuals who move into a sampled household may be mis-matched to other individuals' earlier records. High attrition rates are a particular problem in the linked CPS for young workers. Unless a very large sample size is required, it is often preferable to

[43] See NLS Users' Guide (NLS Handbook, 1995) for further information.

[44] See Polivka (1996) for an analysis of recent changes in the CPS, and for a list of supplements.

use a dataset that was designed to track respondents longitudinally, instead of a linked CPS.

The PSID is a national probability sample that originally consisted of 5000 families in 1968.[45] The original families, and new households that have grown out of those in the original sample, have been followed each year since. Consequently, the PSID provides a unique dataset for studying family-related issues. The number of individuals covered by the PSID increased from 18,000 in 1968 to a cumulative total exceeding 40,000 in 1996, and the number of families increased to nearly 8000. Brown et al. (1996) note that the "central focus of the data is economic and demographic, with substantial detail on income sources and amounts, employment, family composition changes and residential location." The PSID is also one of the few datasets that contains information on consumption and wealth. A recent paper by Fitzgerald et al. (1998) finds that, despite attrition of nearly half the sample since 1968, the PSID remained roughly representative through 1989.[46]

The accessibility of secondary datasets is changing rapidly. The ICPSR remains a major collector and distributor of datasets and codebooks. In addition, CPS data can be obtained directly from the Bureau of Labor Statistics. Increasingly, data collection agencies are making their data directly available to researchers via the internet. In 1996, for example, the Census Bureau made the recent March Current Population Surveys, which include supplemental information on annual income and demographic characteristics, available over the internet. Because the March CPS contains annual income data, many researchers have matched these data from one year to the next.

Because secondary datasets are typically collected for a broad range of purposes or for a purpose other than that intended by the researcher, they often lack information required for a particular project. For example, the PSID would be ideal for a longitudinal study of the impact of personal computers on pay, except it lacks information on the use of personal computers. In other situations, the data collector may omit survey items from public-use files to preserve respondent confidentiality. Nonetheless, several large public-use surveys enable researchers to add questions, or will provide customized extracts with variables that are not on the public-use file. For example, Vroman (1991) added supplemental questions to the CPS on the utilization of unemployment insurance benefits. The cost of adding 7 questions was $100,000.[47] From time to time, survey organizations also solicit researchers' advice on new questions or new modules to add to on-going surveys. Since 1993, for example, the PSID sponsors have held an open competition among researchers to add supplemental questions to the survey.

[45] This paragraph is based on Brown et al. (1996).

[46] See also Becketti et al. (1988) for evidence on the representativeness of the PSID.

[47] Because of concern that the additional questions might affect future responses, the supplement was only asked of individuals who were in their final rotation in the sample. The supplement was added to the survey in the months of May, August, November 1989 and February 1990. The sample size was 2859 eligible unemployed individuals.

3.1.1. Historical comparability in the CPS and Census

Statistical agencies are often faced with a tradeoff between adjusting questions to make them more relevant for the modern economy and maintaining historical comparability. Often it seems that statistical agencies place insufficient weight on historical consistency. For example, after 50 years of measuring education by the highest grade of school individuals attended and completed, the Census Bureau switched to measuring educational attainment by the highest degree attained in the 1990 Census. The CPS followed suit in 1992. This is a subtle change in the education data, but one that could potentially affect estimates of the economic return to education (see Park, 1994; Jaeger, 1993). Because many statistics are most informative in comparison to their values in earlier years, it is important that statistical agencies place weight on historical comparability even though the concepts being measured may have changed.

Fortunately, the Bureau of Labor Statistics and the Census Bureau typically introduce a major change in a questionnaire after studying the likely effects of the change on the survey results. Because some changes have a major impact on certain variables (or on certain populations), it is important that analysts be aware of changes in on-going surveys, and of their likely effects. For example, a major redesign of the CPS was introduced in January 1994, after 8 years of study. The redesigned CPS illustrates the importance of questionnaire changes, as well as the difficulty of estimating the likely impact of such changes.

The redesigned CPS is conducted with computer-assisted interviewing technology, which facilitates more complicated skip patterns, more narrowly tailored questions, and dependent interviewing (in which respondents' answers to an earlier month's question are integrated into the current month's question). In addition, the redesign changed the way key labor force variables were collected in the basic, i.e., non-supplemental, CPS. Most importantly, individuals who are not working are now probed more thoroughly for actions taken to search for work. In the older survey, interviewers were instructed to ask a respondent who "appears to be a homemaker" whether she was keeping house most of last week or doing something else. The new question is gender neutral. Another major change concerns the earnings questions. Prior to the redesign, the CPS asked respondents for their usual weekly wage and usual weekly hours.[48] The ratio of these two variables gives the implied hourly wage. The redesigned CPS first asks respondents for the easiest way they could report their total earnings on their main job (e.g., hourly, weekly, annually, or on some other basis), and then collects usual earnings on that basis.

To gauge the impact of the survey redesign on responses in 1992 and 1993, the BLS and Census Bureau conducted an overlap survey in which a separate sample of households was interviewed using the redesigned CPS, while the regular sample was still given the old CPS questionnaire. Then, for the first 5 months of 1994, this overlap sample was given the old CPS, while the regular sample was given the new one. Overlap samples can be extremely informative, but they are also difficult to implement properly. In this instance,

[48] The old CPS also collected hourly earnings for workers who indicated they were paid hourly.

the overlap sample was drawn with different procedures than the regular CPS sample, and there appear to be systematic differences between the two samples which complicate comparisons. Taking account of these difficulties, Polivka (1996) and Polivka and Miller (1995) estimate that the redesign had an insignificant effect on the unemployment rate, although it appears to have raised the employment-to-population ratio of women by 1.6%, raised the proportion of self-employed women by 20%, increased the proportion of all workers who are classified as part-time by 10%, and decreased the fraction of discouraged workers (i.e., those out of the labor force who have given up searching for work because they believe no jobs are available for them) by 50%. Polivka (1997) addresses the effect of the redesign on the derived hourly wage rate. She finds that the redesign causes about a 5% increase in the average earnings of college graduates relative to those who failed to complete high school, and about a 2% increase in the male-female gap. The potential changes in measurement brought about by the redesigned CPS could lead researchers to incorrectly attribute shifts in employment or wages to economic forces rather than to changes in the questionnaire and survey technology.

Three other changes in the CPS are especially noteworthy. First, beginning in 1980 the Annual Demographic Supplement of the March CPS was expanded to ask a more probing set of income questions. The impact of these changes can be estimated because the 1979 March CPS administered the old (pre-1980) questionnaire to five of the eight rotation groups in the sample, and administered the new, more detailed questionnaire to the other three rotation groups.[49] Second, as noted above, the education question (which is on the "control card" rather than the basic monthly questionnaire) was switched from the number of years of school completed to the highest degree attained in 1992 (see Park, 1994; Jaeger, 1993). Third, the "top code" for the income and earnings questions – that is, the highest level of income reported in the public-use file – has changed over time, which obviously may have implications for studies of income inequality.

3.2. Primary data collection and survey methods

It is increasingly common for labor economists to be involved in collecting their own data. Labor economists' involvement in the design and collection of original datasets takes many forms. First, it should be noted that labor economists have long played a major role in the design and collection of some of the major public-use data files, including the PSID and NLS.

Second, researchers have turned to collecting smaller, customized data to estimate specific quantities or describe certain economic phenomenon. Some of Richard Freeman's research illustrates this approach. Freeman and Hall (1986) conducted a survey to estimate the number of homeless people in the US, which came very close to the official Census

[49] See Krueger (1990a) for an analysis of the change in the questionnaire on responses to the question on workers' compensation benefits. The new questionnaire seems to have detected 20% more workers' compensation recipients. See Coder and Scoon-Rogers (1996) for a comparison of CPS and SIPP income measures.

Bureau estimate in 1990. Borjas et al. (1991) conducted a survey of border crossing behavior of illegal aliens to estimate the number of illegal aliens in the US. Freeman (1990) surveyed inner-city youths in Boston, as part of a follow-up to the survey by Freeman and Holzer (1986). Often, data collected in these surveys are combined with secondary data files to derive national estimates.

Third, some surveys have been conducted to probe the sensitivity of results in large-scale secondary datasets, or to probe the sensitivity of responses to question wording or order. For example, Farber and Krueger (1993) surveyed 102 households in which non-union respondents were asked two different questions concerning their likelihood of joining a union, with the order of the questions randomly interchanged. The two questions, which are listed below, were included in earlier surveys conducted by the Canadian Federation of Labor (CFL) and the American Federation of Labor-Congress of Industrial Organizations (AFL-CIO), and had been analyzed by Riddell (1992). Based on comparing responses to these questions, Riddell concluded that American workers have a higher "frustrated demand" for unions than Canadians:

> CFL Q.: Thinking about your own needs, and your current employment situation and expectations, would you say that it is very likely, somewhat likely, not very likely, or not likely at all that you would consider joining or associating yourself with a union or a professional association in the future?
> AFL Q.: If an election were held tomorrow to decide whether your workplace would be unionized or not, do you think you would definitely vote for a union, probably vote for a union, probably vote against a union, or definitely vote against a union?

In their small-scale survey, Farber and Krueger (1993) found that the responses to the CFL question were extremely sensitive to the questions that preceded them. If the AFL question was asked first, 55% of non-union members answered the CFL question affirmatively, but if the CFL question was asked first, 26% of non-union members answered affirmatively to the CFL question.[50] Thus, the Farber and Krueger results suggest a good deal of caution is warranted when interpreting the CFL-style question, especially across countries.

Finally, and of most interest for our purposes, researchers have conducted special-purpose surveys to evaluate natural experiments or exploit unusual circumstances. Probably the best known example of this type of survey is Card and Krueger's (1994) survey of fast food restaurants in New Jersey and Pennsylvania. Other examples include: Ashenfelter and Krueger's (1994) survey of twins; Behrman et al.'s (1996) survey of twins; Mincer and Higuchi's (1988) survey of turnover at Japanese plants in the US and their self-identified competitors; and Freeman and Kleiner's (1990) survey of companies undergoing a union drive and their competitors.

Several excellent volumes have been written on the design and implementation of

[50] The t-ratio for the difference between the proportions is 3.3.

surveys, and a detailed overview of this material is beyond the scope of this paper.[51] But a few points that may be of special interest to labor economists are outlined below.

Customized surveys seem especially appropriate for rare populations, which are likely to be under-represented or not easily identified in public-use datasets. Examples include identical twins, illegal aliens, homeless people, and disabled people.

To conduct a survey, one must obviously have a questionnaire. Preparing a questionnaire can be a time- consuming and difficult endeavor. Survey researchers often find that answers to questions – even factual economic questions – are sensitive to the wording and ordering of questions. Fortunately, one does not have to begin writing a questionnaire from scratch. Survey questionnaires typically are not copyright protected. Because many economists are familiar with existing questionnaires used in the major secondary datasets (e.g., the CPS), and because a great deal of effort typically goes into designing and testing these questionnaires, it is often advisable to copy as many questions as possible verbatim from existing questionnaires when formulating a new questionnaire. Aside from the credibility gained by replicating questions from well known surveys, another advantage of duplicating others' questions is that the results from the sampled population can be compared directly to the population as a whole with the secondary survey. Furthermore, if data from a customized survey are to be pooled with data from a secondary survey, it is essential that the questions be comparable.

One promising recent development in questionnaire design involves "follow-up brackets" (also known as "unfolding" brackets). This technique offers bracketed categories to respondents who initially refuse or are unable to provide an exact value to an open ended question. Juster and Smith (1997) find that follow-up brackets reduced non-response to wealth questions in the Health and Retirement Survey (HRS) and Asset and Health Dynamics among the Oldest Old Survey (AHEAD). See Hurd, et al. (1998) for experimental evidence of "anchoring effects" in responses based on the sequence of unfolding brackets for consumption and savings data in the AHEAD survey. Follow-up brackets have also been used to measure wealth in the PSID. Follow-up brackets seem particularly useful for hard-to-measure quantities, such as income, wealth, saving and consumption.

Lastly, power calculations should guide the determination of sample size prior to the start of a survey. For example, suppose the goal of the survey is to estimate a 95% confidence interval for a mean. With random sampling, the expected sample size (n) required to obtain a confidence interval of width $2W$ is $n = 4\sigma^2/W^2$, where σ^2 is the population variance of the variable in question. Although the variance generally will not be known prior to conducting the survey, an estimate from other surveys can be used for the power calculation. Also notice that in the case of a binary variable (i.e., if the goal is to estimate a proportion, p), the variance is $p(1-p)$, so in the worse-case scenario the variance is $0.25 = 0.5 \times 0.5$. It should also be noted that in complex sample designs involving clustering and stratification, more observations are usually needed than in simple random samples to attain a given level of precision.

[51] See, e.g., Groves (1989), Sudman and Bradburn (1991), and Singer and Presser (1989).

3.3. Administrative data and record linkage

Administrative data, i.e., data produced as a by-product of some administrative function, often provide inexpensive large samples. The proliferation of computerized record keeping in the last decade should increase the number of administrative datasets available in the future. Examples of widely used administrative data bases include social security earnings records (Ashenfelter and Card, 1985; Vroman, 1990; Angrist, 1990), unemployment insurance payroll and benefit records (Anderson, 1993; Katz and Meyer, 1990; Jacobson et al., 1994; Card and Krueger, 1998), workers' compensation insurance records (Meyer et al., 1995; Krueger, 1990b), company personnel records (Medoff and Abraham, 1980; Lazear, 1992; Baker et al., 1994), and college records (Bowen and Bok, 1998). An advantage of administrative data is that they often contain enormous samples or even an entire population. Another advantage is that administrative data often contain the actual information used to make economic decisions. Thus, administrative data may be particularly useful for identifying causal effects from discrete thresholds in administrative decision making, or for implementing strategies that control for selection on observed characteristics.

A frequent limitation of administrative data, however, is that they may not provide a representative sample of the relevant population. For example, companies that are willing to make their personnel records available are probably not representative of all companies. In some cases administrative data have even been obtained as a by-product of court cases or collected by parties with a vested interest in the outcome of the research, in which case there is additional reason to be concerned about the representativeness of the samples.

Another common limitation of administrative data is that they are not generated with research purposes in mind, so they may lack key variables used in economic analyses. For example, social security earnings records lack data on individuals' education. As a consequence, it is common for researchers to link survey data to administrative data, or to link across administrative datasets. Often these links are based on social security numbers or individuals' names. Examples of linked datasets include: the Continuous Longitudinal Manpower Survey (CLMS) survey, which is a link between social security records and the 1976 CPS; the 1973 Exact Match file which contains CPS, IRS, and social security data; and the Longitudinal Employer-Employee Data Set (LEEDS). All of these linked datasets are now dated, but they can still be used for some important historical studies (e.g., Chay, 1996). More recently, the Census Bureau has been engaged in a project to link Census data to the Survey of Manufacturers.

It is also possible to petition government agencies to release administrative data. Although the Internal Revenue Service severely limits disclosure of federal administrative records collected for tax purposes, State data is often accessible and even federal data can still be linked and released under some circumstances. For example, Angrist (1998) linked military personnel records to Social Security Administration (SSA) data. The HRS has also linked SSA data to survey-based data. Some new Social Security-Census linked datasets are available on a restricted basis through the Census Regional Data Centers. Furthermore, many states provide fairly free access to UI payroll tax data

to researchers for the purpose of linking data.[52] There is also a literature on data release schemes for administrative records that preserve confidentiality and meet legal requirements (see, e.g., Duncan and Pearson, 1991).

3.4. Combining samples

Although in some cases individual records can be linked across different data sources, an alternative linkage strategy exploits the fact that many of the estimators used in empirical research can be constructed from separate sets of first and second moments. So, in principle, individual records with a full complement of variables are not always needed to carry out a multivariate analysis. It is sometimes enough to have all the moments required, even though these moments may be drawn from more than one sample. In practice, this makes it possible to undertake empirical projects even if the required data are not available in any single source.

Recent versions of the multiple-sample approach to empirical work include the two-sample instrumental variables estimators developed by Arellano and Meghir (1992) and Angrist and Krueger (1992, 1995), and used by Lusardi (1996), Japelli et al. (1998), and Kling (1998). The use of two samples to estimate regression coefficients dates back at least to Durbin (1953), who discussed the problem of how to update OLS estimates with information from a new sample. Maddala (1971) discussed a similar problem using a maximum likelihood framework. This idea was recently revived by Imbens and Lancaster (1994), who address the problem of how to use macroeconomic data in micro-econometric models. Deaton (1985) focuses on estimating panel data models with aggregate data on cohorts.

4. Measurement issues

In his classic volume on the accuracy of economic measurement, Morgenstern (1950) quotes the famed mathematician Norbert Wiener as remarking, "Economics is a one or two digit science." The fact that the focus of most empirical research has moved from aggregate time-series data to micro-level cross-sectional and longitudinal survey data in recent years only magnifies the importance of measurement error, because (random) errors tend to average out in aggregate data. Consequently, a good deal of attention has been paid to the extent and impact of "noisy" data in the last decade, and much has been learned.

Measurement error can arise for several reasons. In survey data, a common source of measurement error is that respondents give faulty answers to the questions posed to them.[53] For example, some respondents may intentionally exaggerate their income or

[52] An example is Krueger and Kruse (1996), which links New Jersey unemployment insurance payroll tax data to a dataset the authors collected in a survey of disabled individuals.

[53] Even well-trained economists can make errors of this sort. Harvard's Dean of Faculty Henry Rosovsky (1990, p. 40) gives the following account of a meeting he had with an enraged economics professor who complained about his salary: "After a quick calculation, this quantitatively oriented economist concluded that his raise was all of 1%: an insult and an outrage. I had the malicious pleasure of correcting his mistaken calculation. The raise was 6%: he did not know his own salary and had used the wrong base."

educational attainment to impress the interviewer, or they may shield some of their income from the interviewer because they are concerned the data may somehow fall into the hands of the IRS, or they may unintentionally forget to report some income, or they may misinterpret the question, and so on. Even in surveys like the SIPP, which is specifically designed to measure participation in public programs like UI and AFDC, respondents appear to under-report program participation by 20–40% (see Marquis et al., 1996). It should also be stressed that in many situations, even if all respondents correctly answer the interviewers' questions, the observed data need not correspond to the concept that researchers would like to measure. For example, in principle, human capital should be measured by individuals' acquired knowledge or skills; in practice it is measured by years of schooling.[54]

For these reasons, it is probably best to think of data as routinely being mismeasured. Although few economists consider measurement error the most exciting research topic in economics, it can be of much greater practical significance than several hot issues. Topel (1991), for example, provides evidence that failure to correct for measurement error greatly affects the estimated return to job tenure in panel data models. Fortunately, the direction of biases caused by measurement error can often be predicted. Moreover, in many situations the extent of measurement error can be estimated, and the parameters of interest can be corrected for biases caused by measurement error.

4.1. Measurement error models

4.1.1. The classical model

Suppose we have data on variables denoted X_i and Y_i for a sample of individuals. For example, X_i could be years of schooling and Y_i log earnings. The variables X_i and Y_i may or may not equal the correctly-measured variables the researcher would like to have data on, which we denote X_i^* and Y_i^*. The error in measuring the variables is simply the deviation between the observed variable and the correctly-measured variable: for example, $e_i = X_i - X_i^*$, where e_i is the measurement error in X_i. Considerations of measurement error usually start with the assumption of "classical" measurement errors.[55] Under the classical assumptions, e_i is assumed to have the properties $C(e_i, X_i^*) = E(e_i) = 0$. That is, the measurement error is just mean-zero "white noise". Classical measurement error is not a necessary feature of measurement error; rather, these assumptions are best viewed as a convenient starting point.

What are the implications of classical measurement error? First, consider a situation in which the dependent variable is measured with error. Specifically, suppose that $Y_i = Y_i^* + u_i$, where Y_i is the observed dependent variable, Y_i^* is the correctly-measured,

[54] Measurement error arising from the mismatch between theory and practice also occurs in administrative data. In fact, this may be a more severe problem in administrative data than in survey data.

[55] References for the effect of measurement error include Duncan and Hill (1985), Griliches (1986), Fuller (1987), and Bound and Krueger (1991).

desired, or "true" value of the dependent variable, and u_i is classical measurement error. If Y_i is regressed on one or more correctly-measured explanatory variables, the expected value of the coefficient estimates is not affected by the presence of the measurement error. Classical measurement error in the dependent variable leads to less precise estimates – because the errors will inflate the standard error of the regression – but does not bias the coefficient estimates.[56]

Now consider the more interesting case of measurement error in an explanatory variable. For simplicity, we focus on a bivariate regression, with mean zero variables so we can suppress the intercept. Suppose Y_i^* is regressed on the observed variable X_i, instead of on the correctly-measured variable X_i^*. The population regression of Y_i^* on X_i^* is

$$Y_i^* = X_i^*\beta + \varepsilon_i, \tag{45}$$

while if we make the additional assumption that the measurement error (e_i) and the equation error (ε_i) are uncorrelated, the population regression of Y_i^* on X_i is

$$Y_i^* = X_i\lambda\beta + \tilde{\varepsilon}_i, \tag{46}$$

where $\lambda = C(X^*, X)/V(X)$. If X_i is measured with classical measurement error, then $C(X^*, X) = V(X^*)$ and $V(X) = V(X^*) + V(e)$, so the regression coefficient is necessarily *attenuated*, with the proportional "attenuation bias" equal to $(1 - \lambda) < 1$.[57] The quantity λ is often called the "reliability ratio". If data on both X_i^* and X_i were available, the reliability ratio could be estimated from a regression of X_i^* on X_i. A higher reliability ratio implies that the observed variability in X_i contains less noise.

Although classical measurement error models provide a convenient starting place, in some important situations classical measurement error is impossible. If X_i is a binary variable, for example, then it *must* be the case that measurement errors in X_i are dependent on the values of X_i^*. This is because a dummy variable can only be misclassified in one of two ways (a true 1 can be classified as a 0, and a true 0 can be classified as a 1), so only two values of the error are possible and the error automatically depends on the true value of the variable. An analogous situation arises with variables whose range is limited. Aigner (1973) shows that random misclassification of a binary variable still biases a bivariate regression coefficient toward 0 even though the resulting measurement error is not classical. But, in general, if measurement error in X_i is not classical, the bias factor could be greater than or less than one, depending on the correlation between the measurement error and the true variable. Note, however, that regardless of whether or not the classical

[56] If the measurement error in the dependent variable is not classical, then the regression coefficients will be biased. The bias will equal the coefficients from a hypothetical regression of the measurement error on the explanatory variables.

[57] Notice these are descriptions of population regressions. The estimated regression coefficient is asymptotically biased by a factor $(1 - \lambda)$, although the bias may differ in a finite sample. If the conditional expectation of Y is linear in X, such as in the case of normal errors, the expected value of the bias is $(1 - \lambda)$ in a finite sample.

measurement error assumptions are met, the proportional bias $(1-\lambda)$ is still given by one minus the regression coefficient from a regression of X_i^* on X_i.[58]

Another important special case of non-classical measurement error occurs when a group average is used as a "proxy-variable" for an individual-level variable in micro data. For example, average wages in an industry or county might be substituted for individual wage rates on the right-hand side of an equation if micro wage data are missing. Although this leads to measurement error, since the proxy-variable replaces a desired regressor, asymptotically there is no measurement-error bias in a bivariate regression in this case. One way to see this is to note that the coefficient from a regression of, say, X_i on $E[X_i \mid \text{industry } j]$ has a probability limit of 1.

So far the discussion has considered the case of a bivariate regression with just one explanatory variable. As noted in Section 2, adding additional regressors will typically exacerbate the impact of measurement error on the coefficient of the mismeasured variable because the inclusion of additional independent variables absorbs some of the signal in X_i, and thereby reduces the residual signal-to-noise ratio. Assuming that the other explanatory variables are measured without error, the reliability ratio conditional on other explanatory variables becomes $\lambda' = (\lambda - R^2)/(1 - R^2)$ where R^2 is the coefficient of determination from a regression of the mismeasured X_i on the other explanatory variables. If the measurement error is classical, then $\lambda' \leq \lambda$. And even if the measurement error is not classical, it still remains true that when there are covariates in Eq. (45), the proportional bias is given by the coefficient on X_i in a regression of X_i^* on X_i and the covariates. Note, however, that in models with covariates, the use of aggregate proxy variables may generate asymptotic bias.

An additional feature of measurement error important for applied work is that, for reasons similar to those raised in the discussion of models with covariates, attenuation bias due to classical measurement error is generally exacerbated in panel data models. In particular, if the independent variable is expressed in first differences and if we assume that X_i^* and e_i are covariance stationary, the reliability ratio is

$$\lambda = V(X_i^*)/\{V(X_i^*) + V(e_i)[(1-\tau)/(1-r)]\}, \tag{47}$$

where r is the coefficient of first-order serial correlation in X_i^* and τ is the first-order serial correlation in the measurement error. If the (positive) serial correlation in X_i^* exceeds the (positive) serial correlation in the measurement error, attenuation bias is greater in first-differenced data than in cross-sectional data (Griliches and Hausman, 1986). Classical measurement errors are usually assumed to be serially uncorrelated ($\tau = 0$), in which case the attenuation bias is greater in a first-differenced regression than in a levels regression.

[58] This result requires the previously mentioned assumption that e_i and ε_i be uncorrelated. It may also be the case that the measurement error is not mean zero. Statistical agencies often refer to such phenomenon as "non-sampling error" (see, e.g., McCarthy, 1979). Such non-sampling errors may arise if the questionnaire used to solicit information does not pertain to the economic concept of interest, or if respondents systematically under or over report their answers even if the questions do accurately reflect the relevant economic concepts. An important implication of non-sampling error is that aggregate totals will be biased.

The intuition for this is that some of the signal in X_i cancels out in the first-difference regression because of serial correlation in X_i^*, while the effect of independent measurement errors is amplified because errors can occur in the first or second period. A similar situation arises if differences are taken over dimensions of the data other than time, such as between twins or siblings.

Finally, note that if an explanatory variable is a function of a mismeasured dependent variable, the measurement errors in the dependent and independent variables are automatically correlated. Borjas (1980) notes that this situation often arises in labor supply equations where the dependent variable is hours worked and the independent variable is average hourly earnings, derived by dividing weekly or annual earnings by hours worked. In this situation, measurement error in Y_i will induce a negative bias when $(Y_i^* + u_i)$ is regressed on $X_i^*/(Y_i^* + u_i)$. In other situations, both the dependent and independent variables may have the same noisy measure in the denominator, such as when the variables are scaled to be per capita (common in the economic growth literature). If the true regression parameter were 0, this would bias the estimated coefficient toward 1. The extent of bias in these situations is naturally related to the extent of the measurement error in the variable that appears on both the right-hand and left-hand side of the equation.

4.1.2. Instrumental variables and measurement error

One of the earliest uses of IV was as a technique to overcome errors-in-variables problems. For example, in his classic work on the permanent income hypothesis, Friedman (1957) argued that annual income is a noisy measure of permanent income. The grouped estimator he used to overcome measurement errors in permanent income can be thought of as IV. It is now well known that IV yields consistent parameter estimates even if the endogenous regressor is measured with classical error, assuming that a valid instrument exists. Indeed, one explanation why IV estimates of the return to schooling frequently exceed OLS estimates is that measurement error attenuates the OLS estimates (e.g., Griliches, 1977).

In a recent paper, Kane et al. (1997) emphasize that IV can yield inconsistent parameter estimates if the endogenous regressor is measured with non-classical measurement error.[59] Specifically, they show that if the mismeasured endogenous regressor, X_i, is a dummy variable, the measurement error will be correlated with the instrument, and typically bias the magnitude of IV coefficients upward.[60] The probability limit of the IV estimate in this case is

$$\frac{\beta}{1 - \mathrm{P}(X_i = 0 \mid X_i^* = 1) - \mathrm{P}(X_i = 1 \mid X_i^* = 0)}. \tag{48}$$

Intuitively, the parameter of interest is inflated by one minus the sum of the probabilities of

[59] A similar point has been made by James Heckman in an unpublished comment on Ashenfelter and Krueger (1994).

[60] The exception is if X_i is so poorly measured that it is negatively correlated with X_i^*.

the two types of errors that can be made in measuring X_i (observations that are 1's can be classified as 0's, and observations that are 0's can be classified as 1's). The reason IV tends to overestimate the parameter of interest is that if X_i is a binary variable, the value of the measurement error is automatically dependent on the true value of X_i*, and therefore must be correlated with the instrumental variable because the instrumental variable is correlated with X_i*. Combining this result with the earlier discussion of attenuation bias, it should be clear that if the regressor is a binary variable (in a bivariate regression), the probability limit of the OLS and IV estimators bound the coefficient of interest, assuming the specifications are otherwise appropriate. In the more general case of non-classical measurement error in a continuous explanatory variable, IV estimates can be attenuated or inflated, as in the case of OLS.

4.2. The extent of measurement error in labor data

Mellow and Sider (1983) provide one of the first systematic studies of the properties of measurement error in survey data. They examined two sources of data: (1) employee-reported data from the January 1977 CPS linked to employer-reported data on the same variables for sampled employees; (2) an exact match between employees and employers in the 1980 Employment Opportunity Pilot Project (EOPP). Mellow and Sider focus on the extent of agreement between employer and employee reported data, rather than the reliability of the CPS data per se. For example, they find that 92.3% of employers and employees reported the same one-digit industry, while at the three-digit-industry level, the rate of agreement fell to 71.1%. For wages, they find that the employer-reported data exceeded the employee-reported data by about 5%. The mean unionization rate was slightly higher in the employer-reported data than in the employee-reported data. They also found that estimates of micro-level human capital regressions yielded qualitatively similar results whether employee-reported or employer-reported data are used. This similarity could result from the occurrence of roughly equal amounts of noise in the employer- and employee-reported data.

Several other studies have estimated reliability ratios for key variables of interest to labor economists. Two approaches to estimating reliability ratios have typically been used. First, if the researcher is willing to call one source of data the truth, then λ can be estimated directly as the ratio of the variances: $V(X_i^*)/V(X_i)$. Second, if two measures of the same variable are available (denoted X_{1i} and X_{2i}), and if the errors in these variables are uncorrelated with each other and uncorrelated with the true value, then the covariance between X_{1i} and X_{2i} provides an estimate of $V(X_i^*)$. The reliability ratio λ can then be estimated by using the variance of either measure as the denominator or by using the geometric average of the two variances as the denominator. The former can be calculated as the slope coefficient from a regression of one measure on the other, and the latter can be calculated as the correlation coefficient between the two measures. If a regression approach is used, the variable that corresponds most closely to the data source that is usually used in analysis

should be the explanatory variable (because the two sources may have different error variances).

An example of two mismeasured reports on a single variable are respondents' reports of their parents' education in Ashenfelter and Krueger's (1994) twins study. Each adult twin was asked to report the highest grade of education attained by his or her mother and father. Because each member of a pair of twins has the same parents, the responses should be the same, and there is no reason to prefer one twin's response over the other's. Differences between the two responses for the same pair of twins represent measurement error on the part of at least one twin. The correlation between the twins' reports of their father's education is 0.86, and the correlation between reports of their mother's education is 0.84. These figures probably overestimate the reliability of the parental education data because the reporting errors are likely to be positively correlated; if a parent mis-represented his education to one twin, he is likely to have similarly mis-represented his education to the other twin as well.

Table 9 summarizes selected estimates of the reliability ratio for self-reported log earnings, hours worked, and years of schooling, three of the most commonly studied variables in labor economics. These estimates provide an indication of the extent of attenuation bias when these variables appear as explanatory variables. All of the estimates of the reliability of earnings data in the table are derived by comparing employees' reported earnings data with their employers' personnel records or tax reports. The estimates from the PSID validation study are based on data from a single plant, which probably reduces the variance of correctly-measured variables compared to their variance in the population. This in turn reduces the estimated reliability ratio if reporting errors have the same distribution in the plant as in the population.

Estimates of λ for cross-sectional earnings range from 0.70 to 0.80 for men; λ is somewhat higher for women. The estimated reliability falls to about 0.60 when the earnings data are expressed as year-to-year changes. The decline in the reliability of the earnings data is not as great if 4-year changes are used instead of annual changes, reflecting the fact that there is greater variance in the signal in earnings over longer time periods. Interestingly, the PSID validation study also suggests that hours data are considerably less reliable than earnings data.

The reliability of self-reported education has been estimated by comparing the same individual's reports of his own education at different points in time, or by comparing different siblings' reports of the same individual's education. The estimates of the reliability of education are in the neighborhood of 0.90. Because education is often an explanatory variable of interest in a cross-sectional wage equation, measurement error can be expected to reduce the return to a year of education by about 10% (assuming there are no other covariates). The table also indicates that if differences in educational attainment between pairs of twins or siblings are used to estimate the return to schooling (e.g., Taubman, 1976; Behrman et al., 1980; Ashenfelter and Krueger, 1994; Ashenfelter and Zimmerman, 1997), then the effect of measurement error is greatly exacerbated. This is because schooling levels are highly correlated between twins, while measurement error is

Table 9
Precision of selected variables

	Study	Variable	Dataset	Reliability ratio
1.	Duncan and Hill (1985)	Log earnings 1982	PSID-Validation Study	0.76
		Log earnings 1981		0.71
		Δ Log annual earnings		0.61
2.	Bound and Krueger (1991)	Log annual earnings men	CPS-SER	0.82
		Δ Log annual earnings men		0.65
		Log annual earnings women		0.92
		Δ Log annual earnings women		0.81
3.	Bound et al. (1994)	Log 1986 annual earnings	PSID-Validation Study	0.70
		Log 1982 annual earnings		0.85
		4-year Δ log annual earnings		0.71
		Log 1986 annual hours		0.63
		Log 1982 annual hours		0.72
4.	Siegel and Hodge (1968)	Education	1960 Census Post Enumeration Survey	0.93
5.	Bielby et al. (1977)	Education	1970 Census Post Enumeration Survey	0.80
6.	Ashenfelter and Krueger (1994)	Education	Twinsburg Twins Study	0.90
7.	Behrman et al. (1994)	Education	NAS-NRC Twins Sample	0.92
8.	Behrman et al. (1996)	Education	Minnesota Twin Registry	0.94

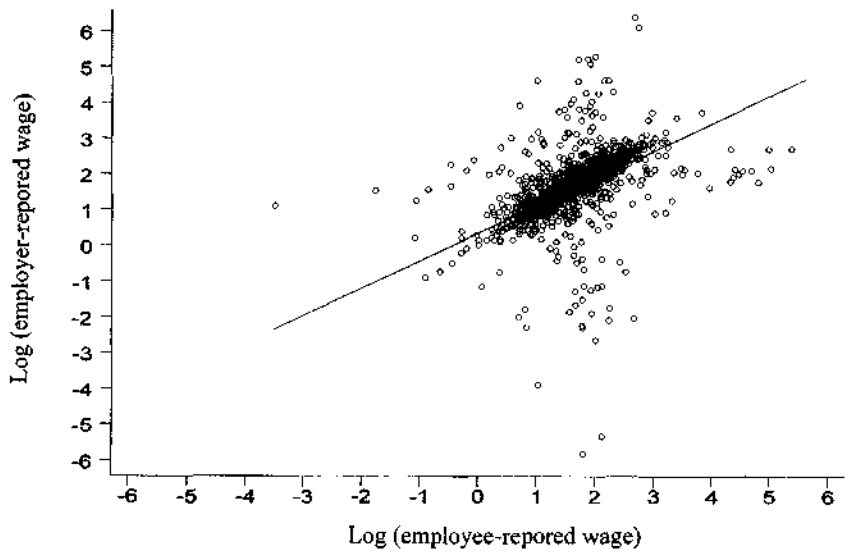

Fig. 6. Scatter plot of employer versus employee-reported log wages, with regression line. Data are from Mellow and Sider (1983).

magnified because reporting errors appear to be uncorrelated between twins. This situation is analogous to the effect of measurement error in panel data models discussed above.

To further explore the extent of measurement error in labor data, we re-analyzed the CPS data originally used by Mellow and Sider (1983). Fig. 6 presents a scatter diagram of the employer-reported log hourly wage against the employee-reported log hourly wage.[61] Although most points cluster around the 45 degree line, there are clearly some outliers. Some of the large outliers probably result from random coding errors, such as a misplaced decimal point.

Researchers have employed a variety of "trimming" techniques to try to minimize the effects of observations that may have been misreported. An interesting study of historical data by Stigler (1977) asks whether statistical methods that downweight outliers would have reduced the bias in estimates of physical constants in 20 early scientific datasets. These constants, such as the speed of light or parallax of the sun, have since been determined with certainty. Of the 11 estimators that he evaluated, Stigler found that the unadjusted sample mean, or a 10% "winsorized mean," provided estimates that were closest to the correct parameters. The 10% winsorized mean sets the values of observations in the

[61] Earnings in the data analyzed by Mellow and Sider were calculated in a manner similar to that used in the redesigned CPS. First, households and firms were asked for the basis on which the employee was paid, and then earnings were collected on that basis. Usual weekly hours were also collected. The household data may have been reported by the worker or by a proxy respondent.

Table 10
Alternative treatment of outliers in Mellow and Sider's matched employee-employer CPS sample[a]

	Mean employee minus employer	r	β	Employee variance	Employer variance
A. Unadjusted data					
ln wage	0.017	0.65	0.77	0.305	0.427
ln hours	−0.043	0.78	0.87	0.147	0.181
B. Employee data winsorized or truncated					
1% winsorized sample					
ln wage	0.021	0.68	0.88	0.258	0.427
ln hours	−0.044	0.77	0.91	0.131	0.181
10% winsorized sample					
ln wage	0.034	0.68	1.04	0.183	0.427
ln hours	−0.069	0.72	1.28	0.057	0.181
1% truncated sample					
ln wage	0.023	0.68	0.91	0.232	0.413
ln hours	−0.041	0.75	0.87	0.117	0.155
10% truncated sample					
ln wage	0.021	0.60	0.94	0.126	0.307
ln hours	−0.030	0.62	0.96	0.031	0.072
C. Both employee and employer data winsorized or truncated					
1% winsorized sample					
ln wage	0.025	0.8	0.86	0.258	0.303
ln hours	−0.04	0.78	0.85	0.131	0.153
10% winsorized sample					
ln wage	0.028	0.88	0.92	0.183	0.198
ln hours	−0.024	0.84	0.85	0.057	0.059
1% truncated sample					
ln wage	0.032	0.88	0.92	0.230	0.250
ln hours	−0.036	0.76	0.81	0.109	0.125
10% truncated sample					
ln wage	0.024	0.91	0.94	0.119	0.125
ln hours	−0.012	0.8	0.83	0.027	0.028

[a] Notes: r is the correlation coefficient between the employee- and employer-reported values. β is the slope coefficient from a regression of the employer-reported value on the employee-reported value. Sample size is 3856 for unadjusted wage data and 3974 for unadjusted hours data. In the 1% winsorized sample, the bottom and top 1% of observations were rolled back to the value corresponding to the 1st or 99th percentile; in the truncated sample these observations were deleted from the sample.

bottom or top decile equal to the value of the observation at the 10th or 90th percentile, and simply calculates the mean for this "adjusted" sample.

In a similar vein, we used Mellow and Sider's linked employer-employee CPS data to explore the effect of various methods for trimming outliers. The analysis here is less clear cut than in Stigler's paper because the true values are not known (i.e., we are not sure the employer-reported data are the "true" data), but we can still compare the reliability of the employee and employer reported data using various trimming methods. The first column of Table 10 reports the difference in mean earnings between the employee and employer responses for the wage and hours data. The differences are small and statistically insignificant. Column 2 reports the correlation between the employee report and the employer report, while column 3 reports the slope coefficient from a bivariate regression of the employer report on the employee report. The regression coefficient in column 3 probably provides the most robust measure of the reliability of the data. Columns 4 and 5 report the variances of the employee and employer data. Results in Panel A are based on the full sample without any trimming. Panel B presents results for a 1% and a 10% "winsorized" sample. We also report results for a 1% and 10% truncated sample. Whereas the winsorized sample rolls back extreme values (defined as the bottom or top $X\%$) but retains them in the sample, the truncated sample simply drops the extreme observations from the sample.[62] In Panel B only the employee-reported data have been trimmed, because that is all that researchers typically observe. In Panel C, we trim both the employee- and employer-reported data.

For hours, the unadjusted data have reliability ratios around 0.80. Interestingly, the reliability of the hours data is considerably higher in Mellow and Sider's data than in the PSID validation study. This may result because the PSID validation study was confined to one plant (which restricted true hours variability compared to the entire workforce), or because there is a difference between the reliability of log weekly hours and annual hours.

The reliability ratio is lower for the wage data than the hours data in the CPS sample. For hours and wages, the correlation coefficients change little when the samples are adjusted (either by winsorizing or truncating the sample), but the slope coefficients are considerably larger in the adjusted data and exceed 1.0 in the 10% winsorized samples. When both the employer and employee data are trimmed, the reliability of the wage data improves considerably, while the reliability of the hours data is not much affected. These results suggest that extreme wage values are likely to be mistakes. Overall, this brief exploration suggests that a small amount of trimming could be beneficial. In a study of the effect of UI benefits on consumption, Gruber (1997) recommends winsorizing the extreme 1% of observations on the dependent variable (consumption), to reduce residual variability. A similar practice seems justifiable for earnings as well.

[62] Loosely speaking, winsorizing the data is desirable if the extreme values are exaggerated versions of the true values, but the true values still lie in the tails. Truncating the sample is more desirable if the extremes are mistakes that bear no resemblance to the true values.

Table 11
Estimates of reliability ratios from Mellow and Sider's CPS dataset[a]

Variable	r	Bivariate β	Multivariate β
ln wage unadjusted	0.65	0.77	0.66
ln wage 1% truncated[b]	0.68	0.91	0.85
ln wage 1% winsorized[b]	0.68	0.88	0.79
ln hours unadjusted	0.78	0.87	0.86
ln hours 1% truncated[b]	0.75	0.87	0.85
ln hours 1% winsorized[b]	0.77	0.91	0.90
Union	0.84	0.84	0.84
2-digit industry premium	0.93	0.93	0.92
1-digit industry premium	0.91	0.92	0.90
1-digit occupation premium	0.84	0.84	0.75

[a] Notes: r is the correlation coefficient between the employee- and employer-reported values. β is the coefficient from a regression of the employer-reported value on the employee-reported value. In the multiple regression, covariates include: highest grade of school completed, high school diploma; college diploma dummy, marital status, non-white, female, potential work experience, potential work experience squared, and veteran status. Sample size varies from 3806 (for industry) to 4087 (for occupation).

[b] Only the employee data were truncated or winsorized.

The estimates in Table 9 or 10 could be used to "inflate" regression coefficients for the effect of measurement error bias, provided that there are no covariates in the equation. Typically, however, regressions include covariates. Consequently, in Table 11 we use Mellow and Sider's CPS sample to regress the employer-reported data on the employee-reported data *and* several commonly used covariates (education, marital status, race, sex, experience and veteran status). For comparison, the first two columns present the correlation coefficient and the slope coefficient from a bivariate regression of the employer on the employee data. The third column reports the coefficient on the employee-reported variable from a multiple regression which specifies the employer-reported variable as the dependent variable, and the corresponding employee-reported variable as an explanatory variable along with other commonly used explanatory variables; this column provides the appropriate estimates of attenuation bias for a multiple regression which includes the same set of explanatory variables as included in the table. Notice that the reliability of the wage data falls from 0.77 to 0.66 once standard human capital controls are included. By contrast, the reliability of the hours data is not very much affected by the presence of control variables, because hours are only weakly correlated with the controls.

Table 11 also reports estimates of the reliability of reported union coverage status, industry and occupation. Assuming the employer-reported data are correct, the bivariate

regression suggests that union status has a reliability ratio of 0.84.[63] Interestingly, this is unchanged when covariates are included. To convert the industry and occupation dummy variables into a one-dimensional variable, we assigned each industry and occupation the wage premium associated with employment in that sector based on Krueger and Summers (1987). The occupation data seem especially noisy, with an estimated reliability ratio of.75 conditional on the covariates.

Earlier we mentioned that classical measurement error has a greater effect if variables are expressed as changes. Although we cannot examine longitudinal changes with Mellow and Sider's data, a dramatic illustration of the effect of measurement error on industry and occupation changes is provided by the 1994 CPS redesign. The redesigned CPS prompts respondents who were interviewed the previous month with the name of the employer that they reported working for the previous month, and then asks whether they still work for that employer. If respondents answer "no," they are asked an independent set of industry and occupation questions. If they answer "yes," they are asked if the usual activities and duties on their job changed since last month. If they report that their activities and duties were unchanged, they are then asked to verify the previous month's description of their occupation and activities. Lastly, if they answer that their activities and duties changed, they are asked an independent set of questions on occupation, activities, and class of worker. Based on pre-tests of the redesigned CPS in 1991, Rothgeb and Cohany (1992) find that the proportion of workers who appear to change three-digit occupations from one month to the next falls from 39% in the old version of the CPS to 7% in the redesigned version.[64] The proportion who change three-digit industry between adjacent months falls from 23% to 5%. These large changes in the gross industry and occupation flows obviously change one's impression of the labor market.[65]

[63] Union status is a dummy variable, so measurement errors will be correlated with true union status. But if union status is correctly reported by employers, the regression coefficient in Table 11 nonetheless provides a consistent estimate of the attenuation bias. Additionally, note that the reliability of data on union status depends on the true fraction of workers who are covered by a union contract. Since union coverage as a fraction of the workforce has declined over time, the reliability ratio might be even lower today. As an extreme example, note that even if the true union coverage rate falls to zero, the measured rate will exceed zero because some (probably around 3%) non-union workers will be erroneously classified as covered by a union. See Freeman (1984), Jakubson (1986) and Card (1996) for analyses of the effect of measurement error in union status in longitudinal data.

[64] It is also possible that dependent interviewing reduces occupational changes because some respondents find it easier to complete the interview by reporting that they did not change employers even if they did. Although this is possible, Rothgeb and Cohany point out that asking independent occupation and industry questions of individuals who report changing employers could result in spurious industry and occupation changes. In addition, the large number of mismatches between employer and employee reported occupation and industry data in Mellow and Sider's dataset are consistent with a finding of grossly overestimated industry and occupation flows.

[65] See also Poterba and Summers (1986), who estimate the measurement error in employment-status transitions.

4.3. Weighting and allocated values

Many datasets use complicated sampling designs and come with sampling weights that reflect the design. Researchers are often confronted with the question of whether to employ sample weights in their statistical analyses to adjust for non-random sampling. For example, if the sampling design uses stratified sampling by state, with smaller states sampled at a higher rate than larger states, then observations from small states should get less weight if national statistics are to be representative. In addition to providing sample weights for this purpose, the Census Bureau also "allocates" answers for individuals who do not respond to a question in one of their surveys. Missing data are allocated by inserting information for a randomly chosen person who is matched to the person with missing data on the basis of major demographic characteristics. Consequently, there are no "missing values" on Census Bureau micro data files. But researchers may decide to include or exclude observations with allocated responses since information that has been allocated is identified with "allocation flags." Unfortunately, although there is a large literature on weighting and survey non-response, this literature has not produced any easy answers that apply to all datasets and research questions (see, e.g., Rubin, 1983; Dickens, 1985; Lillard et al., 1986; Deaton, 1995, 1997; Groves, 1998).[66]

Two datasets where both weighting and allocation issues come up are the CPS and the 1990 Census Public Use Micro Sample (PUMS), neither of which is a simple random sample. The CPS uses a complicated multi-stage probability sample that over-samples some states, and recently oversamples Hispanics in the March survey (see, e.g., US Bureau of the Census, 1992). The 1990 PUMS also deviates from random sampling because of over-sampling of small areas and Native Americans (US Bureau of the Census, 1996).[67] And even random samples may fail to be representative by chance, or because some sampled households are not actually interviewed. The sampling weights including with CPS and PUMS micro data are meant to correct for these features of the sample design, as well as deviations from random sampling due to chance or non-response that affect the age, Sex, Hispanic origin, or race make-up of the sample. Missing data for respondents in these datasets are also allocated. And in the CPS, if someone fails to answer a monthly supplement (e.g., the March income supplement), then entire record is allocated by drawing a randomly matched "donor record" from someone who did respond.

To assess the consequences of weighting and allocation for one important area of research, we estimated a standard human capital earnings function with data from the 1990 March CPS and 1990 5% PUMS for the four permutations of weighting or not weighting, and including or excluding observations with allocated responses. The samples

[66] But see DuMouchel and Duncan (1983), who note that if the object of regression is a MMSE linear approximation to the CEF then estimates from non-random samples should be weighted.

[67] The 1980 PUMS are simple random samples. The CPS was stratified but self-weighting (i.e. all observations were equally likely to be sampled) until January 1978.

Table 12
Weighting and allocation in the Census and CPS[a]

Covariate	1990 Census				March 1990 CPS			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log wages mean	6.405	6.415	6.425	6.437	6.34	6.348	6.351	6.357
Standard deviation	0.746	0.747	0.723	0.721	0.732	0.734	0.717	0.723
Education	0.10932	0.10828	0.10920	0.10813	0.10839	0.11139	0.10950	0.11314
	(0.00047)	(0.00047)	(0.00049)	(0.00049)	(0.00442)	(0.00438)	(0.00459)	(0.00459)
White	0.208	0.213	0.199	0.202	0.194	0.219	0.196	0.211
	(0.003)	(0.003)	(0.004)	(0.003)	(0.030)	(0.027)	(0.031)	(0.029)
Married	0.386	0.387	0.381	0.382	0.386	0.387	0.343	0.362
	(0.004)	(0.003)	(0.004)	(0.004)	(0.031)	(0.029)	(0.032)	(0.031)
Widowed	0.181	0.165	0.190	0.171	0.110	0.200	0.077	0.075
	(0.013)	(0.013)	(0.014)	(0.014)	(0.108)	(0.105)	(0.117)	(0.115)
Divorced or separated	0.193	0.187	0.202	0.196	0.167	0.135	0.141	0.123
	(0.004)	(0.004)	(0.005)	(0.004)	(0.037)	(0.035)	(0.039)	(0.037)
Hispanic	−0.142	−0.151	−0.138	−0.145	−0.125	−0.179	−0.107	−0.155
	(0.005)	(0.005)	(0.005)	(0.005)	(0.040)	(0.048)	(0.041)	(0.049)
Veteran	−0.012	−0.014	−0.018	−0.021	−0.0001	−0.012	−0.002	−0.015
	(0.002)	(0.002)	(0.002)	(0.002)	(0.016)	(0.016)	(0.017)	(0.017)
Potential experience	0.040	0.041	0.041	0.042	0.0005	−0.002	0.013	0.013
	(0.002)	(0.002)	(0.002)	(0.002)	(0.021)	(0.022)	(0.022)	(0.023)
Pot. experience squared*100	−0.055	−0.055	−0.057	−0.057	0.024	0.035	0.003	0.008
	(0.004)	(0.004)	(0.005)	(0.005)	(0.043)	(0.043)	(0.045)	(0.045)
Allocated	Yes	Yes	No	No	Yes	Yes	No	No
Weighted	No	Yes	No	Yes	No	Yes	No	Yes
N	603763	603731	527095	527071	7134	7134	6361	6361

[a] Notes: The table reports OLS estimates of wage equations with the indicated covariates. Standard errors are reported in parentheses. The samples include black and white men aged 40–49 with at least 8 years of schooling. The Census sample excludes active-duty military personnel and the CPS sample excludes military personnel and the Hispanic over-sample. The CPS schooling variable is highest year completed while the census variable is imputed as described in the appendix.

consist of white and black men age 40–49 with at least 8 years of education.[68] Regression results and mean log weekly earnings are summarized in Table 12. In both datasets, the estimated regression coefficients are remarkably similar regardless of whether the equation is estimated by OLS or weighted least squares to adjust for sample weights, and regardless of whether the observations with allocated values are excluded or included in the sample. Moreover, except for potential experience, the regression coefficients are quite similar if they are estimated with either the Census or CPS sample. One notable difference between the datasets, however, is that mean log earnings are about 6 points higher in the Census than the CPS for this age group.

The results in Table 12 suggest that estimates of a human capital earnings function using CPS and Census data are largely insensitive to whether or not the sample is weighted to account for the sample design, and whether or not observations with allocated values are included in the sample. At least for this application, non-random sampling and the allocation of missing values are not very important.[69] It should be noted, however, that the Census Bureau surveys analyzed here are relatively close to random samples, and that the sample strata involve covariates that are included in the regression models. Some of the datasets discussed earlier, most notably the NLSY and the PSID, include large non-random sub-samples that more extensively select or over-sample certain groups using a wider range of characteristics, including racial minorities, low-income respondents, or military personnel. When working with these data is it important to check whether the use of a non-representative sample affects empirical results. Moreover, since researchers often compare results across samples, weighting may be desirable to reduce the likelihood that differences in sample design generate different results.

5. Summary

This chapter attempts to provide an overview of the empirical strategies used in modern labor economics. The first step is to specify a causal question, which we think of as comparing actual and counterfactual states. The next step is to devise a strategy that can, in principle, answer the question. A critical issue in this context is how the causal effect of interest is identified by the statistical analysis. In particular, why does the explanatory variable of interest vary when other variables are held constant? Who is implicitly being compared to whom? Does the source of variation used to identify the key parameters provide plausible "counterfactuals"? And can the identification strategy be tested in a situation in which the causal variable is not expected to have an effect? Finally, imple-

[68] In addition, to make the samples comparable, the Census sample excludes men who were on active duty in the military, and the CPS sample excludes the Hispanic oversample and men in the armed forces. The education variable in both datasets was converted to linear years of schooling based on highest degree attained.

[69] Of course, the standard errors of the estimates should reflect the sample design and account for changes in variability due to allocation. But for samples of this size, the standard errors are extraordinarily small, so adjusting them for these features of the data is probably of second-order importance.

mentation of the empirical strategy requires appropriate data, and careful attention to the many measurement problems that are likely to arise along the way.

Appendix A

A.1. Derivation of Eq. (9) in the text

The model is

$$Y_i = \beta_0 + \rho S_i + \eta_i, \qquad \mathrm{E}[S_i \eta_i] = 0,$$

$$A_i = \gamma_0 + \gamma_1 S_i + \eta_{1i}, \qquad \mathrm{E}[S_i \eta_{1i}] = 0.$$

The coefficient on S_i in a regression of Y_i on S_i and A_i is $C(Y_i, S_{Ai})/V(S_{Ai})$ where

$$S_{Ai} = S_i - \pi_0 - \pi_1 A_i \qquad \text{and} \qquad \pi_1 = \gamma_1 V(S_i)/V(A_i).$$

Also

$$V(S_{Ai}) = V(S_i) - \pi_1^2 V(A_i) = [V(S_i)/V(A_i)][V(A_i) - \gamma_1^2 V(S_i)] = [V(S_i)/V(A_i)]V(\eta_{1i}).$$

So

$$C(Y_i, S_{Ai})/V(S_{Ai}) = \rho + C(\eta_i, S_i - \pi_0 - \pi_1 A_i)/V(S_{Ai}) = \rho - \pi_1 C(\eta_i, A_i)/V(S_{Ai})$$

$$= \rho - \pi_1 C(\eta_i, \eta_{1i})/V(S_{Ai}) = \rho - \gamma_1 \varphi_{01}.$$

A.2. Derivation of Eq. (34) in the text

To economize on notation, we use $\mathrm{E}[Y \mid X, j]$ as shorthand for $\mathrm{E}[Y_i \mid X_i, S_i = j]$. Repeating Eq. (31) in the text without "i" subscripts:

$$\rho_r = \mathrm{E}[Y(S - \mathrm{E}[S \mid X])]/\mathrm{E}[S(S - \mathrm{E}[S \mid X])]$$

$$= \mathrm{E}[E(Y \mid S, X)(S - \mathrm{E}[S \mid X])]/\mathrm{E}[S(S - \mathrm{E}[S \mid X])]. \tag{A.1}$$

Now write

$$\mathrm{E}[Y \mid X, S] = \mathrm{E}[Y \mid X, 0] + \sum_{j=1}^{S} \{\mathrm{E}[Y \mid X, j] - \mathrm{E}[Y \mid X, j-1]\} = \mathrm{E}[Y \mid X, S = 0] + \sum_{j=1}^{S} \rho_{jx}, \tag{A.2}$$

where

$$\rho_{jx} \equiv \mathrm{E}[Y \mid X, j] - \mathrm{E}[Y \mid X, j-1].$$

We first simplify the numerator of ρ_r. Substituting (A.2) into (A.1):

$$\mathrm{E}[E(Y \mid X,S)(S - \mathrm{E}[S \mid X])] = \mathrm{E}\left\{\left(\sum_{j=1}^{\bar{S}} \rho_{jx}\right)(S - \mathrm{E}[S \mid X])\right\}$$

$$= \mathrm{E}\left\{\mathrm{E}\left[\sum_{j=1}^{\bar{S}} \rho_{jx}(S - \mathrm{E}[S \mid X]) \mid X\right]\right\}.$$

Working with the inner expectation,

$$E\left[\sum_{j=1}^{\bar{S}} \rho_{jx}(S - \mathrm{E}[S \mid X]) \mid X\right] = \sum_{s=1}^{\bar{s}} \sum_{j=1}^{s} \rho_{jx}(s - \mathrm{E}[S \mid X])P_{sx},$$

where

$P_{sx} = P(S = s \mid X)$.

Reversing the order of summation, this equals

$$\sum_{j=1}^{\bar{s}} \rho_{jx}\left[\sum_{s=j}^{\bar{s}} (s - \mathrm{E}[S \mid X])P_{sx}\right] = \sum_{j=1}^{\bar{s}} \rho_{jx}\mu_{jx},$$

where

$$\mu_{jx} = \sum_{s=j}^{\bar{s}} (s - \mathrm{E}[S \mid X])P_{sx}.$$

Now, simplifying,

$$\mu_{jx} = \sum_{j=1}^{\bar{s}} sP_{sx} - \sum_{s=j}^{\bar{s}} \mathrm{E}[S \mid X])P_{sx} = (\mathrm{E}[S \mid X, S \geq j] - \mathrm{E}[S \mid X])P(S \geq j \mid X),$$

Since

$$\mathrm{E}[S \mid X] = \mathrm{E}[S \mid X, S \geq j]P(S \geq j \mid X) + \mathrm{E}[S \mid X, S < j](1 - P(S \geq j \mid X)),$$

$$\mu_{jx} = (\mathrm{E}[S \mid S \geq j, X] - \mathrm{E}[S \mid S < j, X])P(S \geq j \mid X)(1 - P(S \geq j \mid X)).$$

So we have shown

$$\mathrm{E}[Y_i(S_i - \mathrm{E}[S_i \mid X_i])] = E\left[\sum_{j=1}^{\bar{s}} \rho_{jx}\mu_{jx}\right].$$

A similar argument for the denominator shows

$$\mathrm{E}[S_i(S_i - \mathrm{E}[S_i \mid X_i])] = E\left[\sum_{j=1}^{\bar{s}} \mu_{jx}\right].$$

Substitute S for j in the summations to get Eq. (34) using the notation in the text.

A.3. Schooling in the 1990 Census

Years of schooling was coded from the 1990 Census categorical schooling variables as follows:

Years of schooling	Educational attainment
8	5th, 6th, 7th, or 8th grade
9	9th grade
10	10th grade
11	11th grade or 12th grade, no diploma
12	High school graduate, diploma or GED
13	Some college, but no degree
14	Completed associate degree in college, occupational program
15	Completed associate degree in college, academic program
16	Completed bachelor's degree, not attending school
17	Completed bachelor's degree, but now enrolled
18	Completed master's degree
19	Completed professional degree
20	Completed doctorate

References

Abadie, Alberto (1998), "Semiparametric estimation of instrumental variable models for causal effects", Mimeo. (Department of Economics, MIT).

Abowd, John M. and Henry S. Farber (1982), "Job queues and the union status of workers", Industrial and Labor Relations Review 35: 354–367.

Aigner, Dennis J. (1973), "Regression with a binary independent variable subject to errors of observation", Journal of Econometrics 1 (1): 49–59.

Altonji, Joseph G. (1986), "Intertemporal substitution in labor supply: evidence from micro data", Journal of Political Economy 94 (3): S176–S215.

Altonji, Joseph G. and Lewis M. Segal (1996), "Small-sample bias in GMM estimation of covariance structures", Journal of Business and Economic Statistics 14 (3): 353–366.

Anderson, Patricia M. (1993), "Linear adjustment costs and seasonal labor demand: evidence from retail trade firms", Quarterly Journal of Economics 108 (4): 1015–1042.

Anderson, Patricia M. and Bruce D. Meyer (1994), "The extent and consequences of job turnover", Brookings Papers on Economic Activity: Microeconomics: 177–236.

Anderson, T.W., Naoto Kunitomo and Takamitsu Sawa (1982), "Evaluation of the distribution function of the limited information maximum likelihood estimator", Econometrica 50: 1009–1027.
Angrist, Joshua D. (1990), "Lifetime earnings and the vietnam era draft lottery: evidence from social security administrative records", American Economic Review 80: 313–335.
Angrist, Joshua D. (1995a), "Introduction to the JBES symposium on program and policy evaluation", Journal of Business and Economic Statistics 13 (2): 133–136.
Angrist, Joshua D. (1995b), "The economic returns to schooling in the West Bank and Gaza strip", American Economic Review 85 (5): 1065–1087.
Angrist, Joshua D. (1998), "Estimating the labor market impact of voluntary military service using social security data on military applicants", Econometrica 66 (2): 249–288.
Angrist, Joshua D. and William N. Evans (1998), "Children and their parents' labor supply: evidence from exogenous variation in family size", American Economic Review, in press.
Angrist, Joshua D. and Guido W. Imbens (1991), "Sources of identifying information in evaluation models", Technical working paper no. 117 (NBER, Cambridge, MA).
Angrist, Joshua D. and Guido W. Imbens (1995), "Two-stage least squares estimates of average causal effects in models with variable treatment intensity", Journal of the American Statistical Association 90 (430): 431–442.
Angrist, Joshua D. and Alan B. Krueger (1991), "Does compulsory school attendance affect schooling and earnings?" Quarterly Journal of Economics 106: 979–1014.
Angrist, Joshua D. and Alan B. Krueger (1992), "The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples", Journal of the American Statistical Association 87 (418): 328–336.
Angrist, Joshua D. and Alan B. Krueger (1995), "Split-sample instrumental variables estimates of the returns to schooling", Journal of Business and Economic Statistics 13 (2): 225–235.
Angrist, Joshua D. and Victor Lavy (1998), "Using maimonides rule to estimate the effects of class size on scholastic achievement", Quarterly Journal of Economics, in press.
Angrist, Joshua D. and Whitney K. Newey (1991), "Over-identification tests in earnings functions with fixed effects", Journal of Business and Economic Statistics 9 (3): 317–323.
Angrist, Joshua D., Guido W. Imbens and Kathryn Graddy (1995), "Non-parametric demand analysis with an application to the demand for fish", Technical working paper no. 178 (NBER, Cambridge, MA).
Angrist, Joshua D., Guido W. Imbens and Donald B. Rubin (1996), "Identification of causal effects using instrumental variables", Journal of the American Statistical Association 91 (434): 444–455.
Angrist, Joshua D., Guido W. Imbens and Alan B. Krueger (1998), "Jackknife instrumental variables estimation", Journal of Applied Econometrics, in press.
Arellano, Manuel and Costas Meghir (1992), "Female labour supply and on-the-job search: an empirical model estimated using complementary datasets", Review of Economic Studies 59 (3): 537–559.
Ashenfelter, Orley A. (1978), "Estimating the effect of training programs on earnings", Review of Economics and Statistics 60 (1): 47–57.
Ashenfelter, Orley A. (1984), "Macroeconomic analyses and microeconomic analyses of labor supply", Carnegie-Rochester Series on Public Policy 21: 117–155.
Ashenfelter, Orley A. and David E. Card (1985), "Using the longitudinal structure of earnings to estimate the effect of training programs", Review of Economics and Statistics 67 (4): 648–660.
Ashenfelter, Orley A. and Alan B. Krueger (1994), "Estimates of the economic return to schooling from a new sample of twins", American Economic Review 84 (5): 1157–1173.
Ashenfelter, Orley A. and Joseph D. Mooney (1968), "Graduate education, ability and earnings", Review of Economics and Statistics 50 (1): 78–86.
Ashenfelter, Orley A. and David J. Zimmerman (1997), "Estimates of the returns to schooling from sibling data: fathers, sons and brothers", Review of Economics and Statistics 79 (1): 1–9.
Baker, George, Michael Gibbs and Bengt Holmstrom (1994), "The internal economics of the firm: evidence from personnel data", Quarterly Journal of Economics 109 (4): 881–919.

Barnow, Burt S., Glen G. Cain and Arthur Goldberger (1981), "Selection on observables", Evaluation Studies Review Annual 5: 43–59.

Becketti, Sean, William Gould, Lee Lillard and Finis Welch (1988), "The panel study of income dynamics after fourteen years: an evaluation", Journal of Labor Economics 6 (4): 472–492.

Behrman, Jere, Zdenek Hrubec, Paul Taubman and Terence Wales (1980), Socioeconomic success: a study of the effects of genetic endowments, family environment and schooling (North-Holland, Amsterdam).

Behrman, Jere R., Mark R. Rosenzweig and Paul Taubman (1994), "Endowments and the allocation of schooling in the family and in the marriage market: the twins experiment", Journal of Political Economy 102 (6): 1131–1174.

Behrman, Jere R., Mark R. Rosenzweig and Paul Taubman (1996), "College choice and wages: estimates using data on female twins", Review of Economics and Statistics 78 (4): 672–685.

Bekker, Paul A. (1994), "Alternative approximations to the distributions of instrumental variables estimators", Econometrica 62 (3): 657–681.

Bielby, William, Robert Hauser and David Featherman (1977), "Response errors of non-black males in models of the stratification process", in: D.J. Aigner and A.S. Goldberger, eds., Latent variables in socioeconomic models (North-Holland, Amsterdam) pp. 227–251.

Björklund, Anders and Robert Moffitt (1987), "The estimation of wage gains and welfare gains in self-selection models", The Review of Economics and Statistics 69 (1): 42–49.

Borjas, George J. (1980), "The relationship between wages and weekly hours of work: the role of division bias", Journal of Human Resources 15 (3): 409–423.

Borjas, George J., Richard B. Freeman and Kevin Lang (1991), "Undocumented Mexican-born workers in the United States: how many, how permanent?" in: John M. Abowd and Richard B. Freeman, eds., Immigration, trade and the labor market (National Bureau of Economic Research Project Report, University of Chicago Press, Chicago, IL).

Borjas, George J., Richard B. Freeman and Lawrence F. Katz (1997), " How much do immigration and trade affect labor market outcomes?" Brookings Papers on Economic Activity 10 (1): 1–67.

Bound, John (1989), "The health and earnings of rejected disability insurance applicants", American Economic Review 79 (3): 482–503.

Bound, John (1991), "The health and earnings of rejected disability insurance applicants: reply", American Economic Review 81 (5): 1427–1434.

Bound, John and Alan B. Krueger (1991), "The extent of measurement error in longitudinal earnings data: do two wrongs make a right?" Journal of Labor Economics 9 (1): 1–24.

Bound, John, et al. (1994), "Evidence on the validity of cross-sectional and longitudinal labor market data", Journal of Labor Economics 12 (3): 345–368.

Bound, John, David Jaeger and Regina Baker (1995), "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak", Journal of the American Statistical Association 90 (430): 443–450.

Bowen, William G. and Derek Bok (1998), The shape of the river: long-term consequences of considering race in college and university admissions (Princeton University Press, Princeton, NJ).

Bronars, Stephen G. and Jeff Grogger (1994), "The economic consequences of unwed motherhood: using twins as a natural experiment", American Economic Review 84 (5): 1141–1156.

Brown, Charles, Greg J. Duncan and Frank P. Stafford (1996), "Data watch: the panel study of income dynamics", Journal of Economic Perspectives 10 (2): 155–168.

Burtless, Gary (1995), "The case for randomized field trials in economic and policy research", Journal of Economic Perspectives 9 (2): 63–84.

Buse, A. (1992), "The bias of instrumental variable estimators", Econometrica 60 (1): 173–180.

Campbell, Donald T. (1969), "Reforms as experiments", American Psychologist XXIV: 409–429.

Campbell, Donald T. and J.C. Stanley (1963), Experimental and quasi-experimental designs for research (Rand-McNally, Chicago, IL).

Card, David E. (1989), "The impact of the Mariel boatlift on the Miami labor market", Working paper no. 253 (Industrial Relations Section, Princeton University).

Card, David E. (1990), "The impact of the Mariel boatlift on the Miami labor market", Industrial and Labor Relations Review 43: 245–257.

Card, David E. (1995), "Earnings, schooling and ability revisited", in: Solomon W. Polachek, ed., Research in labor economics (JAI Press, Greenwich, CT).

Card, David E. (1996), "The effect of unions on the structure of wages: a longitudinal analysis", Econometrica 64 (4): 957–979.

Card, David E. and Alan B. Krueger (1994), "Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania", American Economic Review 84 (4): 772–784.

Card, David E. and Alan B. Krueger (1998), "A reanalysis of the effect of the New Jersey minimum wage increase on the fast-food industry with representative payroll data", Working paper no. 6386 (NBER, Cambridge, MA).

Card, David E. and Daniel Sullivan (1988), "Measuring the effect of subsidized training on movements in and out of employment", Econometrica 56 (3): 497–530.

Center for Drug Evaluation and Research (1988), Guideline for the format and content of the clinical and statistical sections of a new drug application (US Food and Drug Administration, Department of Health and Human Services, Washington, DC).

Chamberlain, Gary (1977), "Education, income and ability revisited", Journal of Econometrics 5 (2): 241–257.

Chamberlain, Gary (1978), "Omitted variables bias in panel data: estimating the returns to schooling", Annales de l'INSEE 30–31: 49–82.

Chamberlain, Gary (1980), "Discussion", American Economic Review 70 (2): 47–49.

Chamberlain, Gary (1984), "Panel data", in: Zvi Griliches and Michael D. Intriligator, eds., Handbook of econometrics (North-Holland, Amsterdam).

Chamberlain, Gary and Guido W. Imbens (1996), "Hierarchical Bayes models with many instrumental variables", Discussion paper no. 1781 (Department of Economics, Harvard University).

Chamberlain, Gary and Edward E. Leamer (1976), "Matrix weighted averages and posterior bounds", Journal of the Royal Statistical Society, Series B 38: 73–84.

Chay, Kenneth Y. (1996), "An empirical analysis of black economic progress over time" PhD Thesis (Department of Economics, Princeton University).

Cochran, William G. (1965), "The planning of observational studies of human populations (with discussion)", Journal of the Royal Statistical Society, Series A 128: 234–266.

Coder, John and Lydia Scoon-Rogers (1996), "Evaluating the quality of income data collected in the annual supplement to the March Current Population Survey and the Survey of Income and Program Participation", Census working paper no. 215 (US Bureau of the Census, Washington, DC).

Dawid, A.P. (1979), "Conditional independence in statistical theory", Journal of the Royal Statistical Society, Series B 41: 1–31.

Deaton, Angus (1985), "Panel data from a time series of cross-sections", Journal of Econometrics 30: 109–126.

Deaton, Angus (1995), "Data and econometric tools for development analysis", in: Hollis Chenery and T.N. Srinivasan, eds., Handbook of development economics (North-Holland, Amsterdam).

Deaton, Angus (1997), The analysis of household surveys: a microeconometric approach to development policy (Johns Hopkins University Press, Baltimore, MD).

Deaton, Angus and Christina Paxson (1998), "Economies of scale, household size and the demand for food", Journal of Political Economy, in press.

Dehejia, Rajeev H. and Sadek Wahba (1995), "Causal effects in nonexperimental studies: re-evaluating the evaluation of training programs", Mimeo. (Department of Economics, Harvard University).

Dickens, William T. (1985), "Error components in grouped data: why it's never worth weighting", Technical working paper no. 43 (NBER, Cambridge, MA).

Dominitz, Jeff and Charles F. Manski (1997), "Using expectations data to study subjective income expectations", Journal of the American Statistical Association 92: 855–867.

Donald, Steven and Whitney K. Newey (1997), "Choosing the number of instruments", Mimeo. (Department of Economics, MIT).

DuMouchel, William H. and Greg Duncan (1983), "Using sample survey weights in multiple regression analyses of stratified samples", Journal of the American Statistical Association 78, 535–543.

Duncan, Greg J. and Daniel H. Hill (1985), "An investigation of the extent and consequences of measurement error in labor-economic survey data", Journal of Labor Economics 3 (4): 508–532.

Duncan, Greg T. and Robert W. Pearson (1991), "Enhancing access to microdata while protecting confidentiality: prospects for the future", Statistical Science 6 (3): 219–239.

Durbin, J. (1953), "A note on regression when there is extraneous information about one of the coefficients", Journal of the American Statistical Association 48: 799–808.

Durbin, J. (1954), "Errors in variables", Review of the International Statistical Institute 22: 23–32.

Farber, Henry S. and Alan B. Krueger (1993), "Union membership in the United States: the decline continues". Working paper no. 306 (Industrial Relations Section, Princeton University).

Fitzgerald, John, Peter Gottschalk and Robert Moffit (1998), "An analysis of sample attrition in panel data: the Michigan panel study of income dynamics", Journal of Human Resources, in press.

Freeman, Richard B. (1984), "Longitudinal analyses of the effects of trade unions", Journal of Labor Economics 2: 1–26.

Freeman, Richard B. (1989), Labor markets in action (Harvard University Press, Cambridge, MA).

Freeman, Richard B. (1990), "Employment and earnings of disadvantaged young men in a labor shortage economy", Working paper no. 3444 (NBER, Cambridge, MA).

Freeman, Richard B. and Brian Hall (1986), "Permanent homelessness in America?", Working paper no. 2013 (NBER, Cambridge, MA).

Freeman, Richard B. and Harry J. Holzer (1986), "The black youth employment crisis: summary of findings", in: Richard B. Freeman and Harry J. Holzer, eds., The black youth employment crisis (National Bureau of Economic Research Project Report, University of Chicago Press, Chicago, IL).

Freeman, Richard B. and Morris M. Kleiner (1990), "The impact of new unionization on wages and working conditions", Journal of Labor Economics 8 (1): S8–S25.

Friedberg, Rachel M. and Jennifer Hunt (1995), "The impact of immigrants on host country wages, employment and growth", Journal of Economic Perspectives 9 (2): 23–44.

Friedman, Milton (1957), A theory of the consumption function (Princeton University Press, Princeton, NJ).

Fuchs, Victor, Alan B. Krueger and James M. Poterba (1998), "Why do economists disagree about policy? The roles of beliefs about parameters and values", Journal of Economic Perspectives, in press.

Fuller, Wayne A. (1987), Measurement error models (Wiley, New York).

Girshick, M.A. and Trygve Haavelmo (1947), "Statistical analysis of the demand for food: examples of simultaneous estimation of structural equations", Econometrica 15 (2): 79–110.

Goldberger, Arthur S. (1972), "Selection bias in evaluating treatment effects: some formal illustrations", Discussion paper (Institute for Research on Poverty, University of Wisconsin) pp. 123–172.

Goldberger, Arthur S. (1991), A course in econometrics (Harvard University Press, Cambridge, MA).

Gorseline, Donald E. (1932), The effect of schooling upon income (University of Indiana, Bloomington, IN).

Griliches, Zvi (1977), "Estimating the returns to schooling: some econometric problems", Econometrica 45 (1): 1–22.

Griliches, Zvi (1979), "Sibling models and data in economics: beginnings of a survey", Journal of Political Economy 87 (5): S37–S64.

Griliches, Zvi (1986), "Economic data issues", in: Zvi Griliches and Michael D. Intriligator, eds., Handbook of econometrics (North-Holland, Amsterdam).

Griliches, Zvi and Jerry A. Hausman (1986), "Errors in variables in panel data", Journal of Econometrics 31 (1): 93–118.

Griliches, Zvi and Jacques Mairesse (1998), "Production functions: the search for identification", in: Zvi Griliches, ed., Practicing econometrics: essays in method and application (Edward Elgar, Cheltenham, UK).

Griliches, Zvi and William M. Mason (1972), "Education, Income and Ability", Journal of Political Economy 80 (3): S74–S103.

Grosh, Margaret E. and Paul Glewwe (1996), "Household survey data from developing countries: progress and prospects", American Economic Review 86 (2): 15–19.

Grosh, Margaret E. and Paul Glewwe (1998), "Data watch: the World Bank's living standards measurement study household surveys", Journal of Economic Perspectives 12 (1): 187–196.

Groves, Robert M. (1989), Survey errors and survey costs (Wiley, New York).

Groves, Robert M. (1998), Non-response in household interview surveys (Wiley, New York).

Gruber, Jonathan (1997), "The consumption smoothing benefits of unemployment insurance", American Economic Review 87 (1): 192–205.

Hahn, Jinyong (1998), "On the role of the propensity score in the efficient estimation of average treatment effects", Econometrica 66: 315–332.

Hahn, Jinyong, Petra Todd and Wilbert van der Klaauw (1998), "Estimation of treatment effects with a quasi-experimental regression-discontinuity design: with application to evaluating the effect of federal antidiscrimination laws on minority employment in small U.S. firms", Mimeo. (Department of Economics, University of Pennsylvania).

Hall, Alastair R., Glenn D. Rudebusch and David W. Wilcox (1996), "Judging instrument relevance in instrumental variables estimation", International Economic Review 37 (2): 283–298.

Hansen, Lars Peter (1982), "Large sample properties of generalized method of moments estimators", Econometrica 50 (4): 1029–1054.

Hansen, W. Lee, Burton A. Weisbrod and William J. Scanlon (1970), "Schooling and earnings of low achievers", American Economic Review 60 (3): 409–418.

Hausman, Jerry A. and William E. Taylor (1981), "Panel data and unobservable individual effects", Econometrica 49 (6): 1377–1398.

Hearst, Norman, Thomas Newman and Steven Hulley (1986), "Delayed effects of the military draft on mortality: a randomized natural experiment", New England Journal of Medicine 314: 620–624.

Heckman, James J. (1978), "Dummy endogenous variables in a simultaneous equations system", Econometrica 46 (4): 931–959.

Heckman, James J. and V. Joseph Hotz (1989), "Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training", Journal of the American Statistical Association 84 (408): 862–874.

Heckman, James J. and Thomas E. MaCurdy (1986), "Labor econometrics", in: Orley Ashenfelter and Richard Layard, eds., Handbook of labor economics (North-Holland, Amsterdam).

Heckman, James J. and Brook S. Payner (1989), "Determining the impact of antidiscrimination policy on the economic status of blacks: a study of South Carolina", American Economic Review 79 (1): 138–177.

Heckman, James J. and Richard Robb, Jr. (1985), "Alternative methods for evaluating the impact of interventions", in: James J. Heckman and Burton Singer, eds., Longitudinal analysis of labor market data, Econometric society monographs series no. 10 (Cambridge University Press, Cambridge, MA).

Heckman, James J. and Jeffrey A. Smith (1995), "Assessing the case for social experiments", Journal of Economic Perspectives 9 (2): 85–110.

Heckman, James J., Hidehiko Ichimura and Petra E. Todd (1997), "Matching as an econometric evaluation estimator: evidence from evaluating a job training programme", Review of Economic Studies 64 (4): 605–654.

Heckman, James J., Lance Lochner and Christopher Taber (1998), "Tax policy and human-capital formation", American Economic Review 88 (2): 293–297.

Holland, Paul W. (1986), "Statistics and causal inference", Journal of the American Statistical Association 81: 945–970.

Hurd, Michael, et al. (1998), "Consumption and savings balances of the elderly: experimental evidence on survey response bias", in: D. Wise (ed.), Frontiers in the economics of aging (University of Chicago Press, Chicago, IL) pp. 353–387.

Imbens, Guido W. and Joshua D. Angrist (1994), "Identification and estimation of local average treatment effects", Econometrica 62 (2): 467–475.

Imbens, Guido W. and Tony Lancaster (1994), "Combining micro and macro data in microeconometric models", Review of Economic Studies 61 (4): 655–680.

Imbens, Guido W. and Wilbert van der Klaauw (1995), "The cost of conscription in the Netherlands", Journal of Business and Economic Statistics 13 (2): 207–215.

Imbens, Guido W., Donald B. Rubin and Bruce I. Sacerdote (1997), "Estimating income effects: evidence from a survey of lottery players", Mimeo. (Economics Department, UCLA).

Jacobson, Louis S., Robert J. Lalonde and Daniel G. Sullivan (1994), "Earnings losses of displaced workers", American Economic Review 83 (4): 685–709.

Jaeger, David (1993), "The new current population survey education variable: a recommendation", Research report no. 93-289 (Population Studies Center, University of Michigan).

Jakubson, George (1986), "Measurement error in binary explanatory variables in panel data models: why do cross section and panel estimates of the union wage effect differ?" Working paper no. 209 (Industrial Relations Section, Princeton University).

Jakubson, George (1991), "Estimation and testing of the union wage effect using panel data", Review of Economic Studies 58 (5): 971–991.

Jappelli, Taillio, Jorn-Steffen Pischke and Nicholas Souleles (1998), "Testing for liquidity constraints in euler equations with complementary data sources", Review of Economics and Statistics 80, 251–262.

Juhn, Chinhui, Kevin M. Murphy and Brooks Pierce (1993), "Wage inequality and the rise in returns to skill", Journal of Political Economy 101 (3): 410–442.

Juster, F. Thomas and James P. Smith (1997), "Improving the quality of economic data: lessons from the HRS and AHEAD", Journal of the American Statistical Association 92 (440): 1268–1278.

Kane, Thomas J., Cecilia Elena Rouse and Douglas Staiger (1997), "Estimating returns to schooling when schooling is misreported", Unpublished paper.

Katz, Lawrence F. and Bruce Meyer (1990), "Unemployment insurance, recall expectations and unemployment outcomes", Quarterly Journal of Economics 105 (4): 973–1002.

Katz, Lawrence F. and Kevin M. Murphy (1992), "Changes in relative wages, 1963–1987: supply and demand factors", Quarterly Journal of Economics 107 (1): 35–78.

Keane, Michael P. and Kenneth Wolpin (1997), "Introduction to the JBES special issue on structural estimation in applied microeconomics", Journal of Business and Economic Statistics 15 (2): 111–114.

Kling, Jeffrey (1998), "Interpreting instrumental variables estimates of the returns to schooling", in: Identifying causal effects of public policies, PhD thesis (Department of Economics, MIT).

Kremer, Michael (1997), "Development datasets", Mimeo. (Department of Economics, MIT).

Krueger, Alan B. (1990a), "Incentive effects of workers' compensation insurance", Journal of Public Economics 41: 73–99.

Krueger, Alan B. (1990b), "Workers' compensation insurance and the duration of workplace injuries", Working paper no. 3253 (NBER, Cambridge, MA).

Krueger, Alan B. and Douglas Kruse (1996), "Labor market effects of spinal cord injuries in the dawn of the computer age", Working paper no. 349 (Industrial Relations Section, Princeton University).

Krueger, Alan B. and Jorn Steffen Pischke (1992), "The effect of social security on labor supply. A cohort analysis of the notch generation", Journal of Labor Economics 10 (2): 412–437.

Krueger, Alan B. and Lawrence H. Summers (1987), "Efficiency wages and the inter-industry wage structure", Econometrica 56 (2): 259–293.

Lalonde, Robert J. (1986), "Evaluating the econometric evaluations of training programs using experimental data", American Economic Review 76 (4): 602–620.

Lang, Kevin (1993), "Ability bias, discount rate bias and the return to education", Mimeo. (Department of Economics, Boston University).

Lazear, Edward P. (1992), "The job as a concept", in: William J. Bruns, Jr., ed., Performance measurement, evaluation and incentives (Harvard Business School Press, Boston, MA).

Leamer, Edward E. (1982), "Let's take the con out of econometrics", American Economic Review 73 (1): 31–43.
Lester, Richard A. (1946), "Shortcomings of marginal analysis for wage-employment problems", American Economic Review 36: 63–82.
Levy, Frank (1987), Dollars and dreams: the changing american income distribution (Russell Sage Foundation, New York).
Lewis, H. Gregg (1963), Unionism and relative wages in the United States: an empirical inquiry (University of Chicago Press, Chicago, IL).
Lewis, H. Gregg (1986), Union relative wage effects (University of Chicago Press, Chicago, IL).
Lillard, Lee, James P. Smith and Finis Welch (1986), "What do we really know about wages? The importance of nonreporting and census imputation", Journal of Political Economy 94 (3): 489–506.
Lusardi, Ann Maria (1996), "Permanent income, current income and consumption: evidence from two panel datasets", Journal of Business and Economic Statistics 14 (1): 81–90.
MaCurdy, Thomas E. (1981), "An empirical model of labor supply in a life-cycle setting", Journal of Political Economy 89 (6): 1059–1085.
Maddala, G.S. (1971) "The likliehood approach to pooling cross-section and time-series data", Econometrica 39: 939–953.
Marquis, K.H., J.C. Moore and K. Bogen (1996), "An experiment to reduce measurement error in the SIPP: preliminary results", Mimeo. (Bureau of the Census).
Marshall, Alfred (1890), Principles of economics (Macmillan, London).
McCarthy, P.J. (1979), "Some sources of error in labor force estimates from the current population survey", in: National Commission on Employment and Unemployment Statistics, Counting the labor force, appendix, Vol. II (US Government Printing Office, Washington, DC).
Medoff, James L. and Katharine G. Abraham (1980), "Experience, performance and earnings", Quarterly Journal of Economics 95 (4): 703–736.
Mellow, Wesley and Hal Sider (1983), "Accuracy of response in labor market surveys: evidence and implications", Journal of Labor Economics 1 (4): 331–344.
Meyer, Bruce D. (1995), "Natural and quasi-experiments in economics", Journal of Business and Economic Statistics 13 (2): 151–161.
Meyer, Bruce D., W. Kip Viscusi and David L. Durbin (1995), "Workers' compensation and injury duration: evidence from a natural experiment", American Economic Review 85: 322–340.
Mincer, Jacob and Yoshio Higuchi (1988), "Wage structures and labor turnover in the U.S. and in Japan", Journal of the Japanese and International Economy 2 (2): 97–133.
Morgenstern, Oskar (1950), On the accuracy of economic observations (Princeton University Press, Princeton, NJ).
Murphy, Kevin M. and F. Welch (1992), "The structure of wages", Quarterly Journal of Economics 107 (1): 285–326.
Newey, Whitney K. (1985), "Generalized method of moments estimation and testing", Journal of Econometrics 29 (3): 229–256.
Nickell, Stephen J. (1981), "Biases in dynamic models with fixed effects", Econometrica 49 (6): 1417–1426.
NLS Handbook (1995), NLS handbook (Center for Human Resource Research, The Ohio State University, Columbus, OH).
Park, Jin Huem (1994), "Returns to schooling: a peculiar deviation from linearity", Working paper no. 339 (Industrial Relations Section, Princeton University).
Parsons, Donald O. (1980), "The decline in male labor force participation", Journal of Political Economy 88 (1): 117–134.
Parsons, Donald O. (1991), "The health and earnings of rejected disability insurance applicants: comment", American Economic Review 81 (5): 1419–1426.
Passell, P. (1992), "Putting the science in social science", New York Times.
Pindyk, Robert S. and Daniel L. Rubinfeld (1991), Econometric models and economic forecasts (McGraw-Hill, New York).

Polivka, Anne (1996), "Data watch: the redesigned current population survey", Journal of Economic Perspectives 10 (3): 169–181.

Polivka, Anne (1997), "Using earnings data from the current population survey after the redesign", Working paper no. 306 (Bureau of Labor Statistics).

Polivka, Anne and Stephen Miller (1995), "The CPS after the redesign: refocusing the economic lens", Mimeo. (Bureau of Labor Statistics).

Poterba, James M. and Lawrence H. Summers (1986), "Reporting errors and labor market dynamics", Econometrica 54 (6): 1319–1338.

Powell, James L., James H. Stock and Thomas M. Stoker (1989), "Semiparametric estimation of index coefficients", Econometrica 57 (6): 1403–1430.

Riddell, W. Craig (1992), "Unionization in Canada and the United States: a tale of two countries", Mimeo. (Department of Economics, University of British Columbia).

Rosenbaum, Paul R. (1984), "The consequences of adjustment for a concomitant variable that has been affected by the treatment", Journal of the Royal Statistical Society Series A 149: 656–666.

Rosenbaum, Paul R. (1995), Observational studies (Springer-Verlag, New York).

Rosenbaum, Paul R. and Donald B. Rubin (1983), "The central role of the propensity score in observational studies for causal effects", Biometrika 70: 41–55.

Rosenbaum, Paul R. and Donald B. Rubin (1984), "Reducing bias in observational studies using subclassification on the propensity score", Journal of the American Statistical Association 79: 516–524.

Rosenbaum, Paul R. and Donald B. Rubin (1985), "Constructing a control group using multi-variate matching methods that include the propensity score", American Statistician 39: 33–38.

Rosenzweig, Mark R. and Kenneth I. Wolpin (1980), "Testing the quantity-quality model of fertility: the use of twins as a natural experiment", Econometrica 48 (1): 227–240.

Rosovsky, Henry (1990), The university: an owner's manual (W.W. Norton and Company, New York).

Rothgeb, Jennifer M. and Sharon R. Cohany (1992), "The revised CPS questionnaire: differences between the current and the proposed questionnaires", Paper presented at the Annual Meeting of the American Statistical Association.

Rubin, Donald B. (1973), "Matching to remove bias in observational studies", Biometrics 29 (1): 159–183.

Rubin, Donald B. (1974), "Estimating causal effects of treatments in randomized and non-randomized studies", Journal of Educational Psychology 66: 688–701.

Rubin, Donald B. (1977), "Assignment to a treatment group on the basis of a covariate", Journal of Educational Statistics 2: 1–26.

Rubin, Donald B. (1983), "Imputing income in the CPS: comments on 'measures of aggregate labor cost in the United States'", in: Jack E. Triplett, ed., The measurement of labor cost (University of Chicago Press, Chicago, IL).

Sawa, Takamitsu (1969), "The exact sampling distribution of ordinary least squares and two-stage least squares estimators", Journal of the American Statistical Association 64 (327): 923–937.

Sawa, Takamitsu (1973), "An amost unbiased estimator in simultaneous equations systems", International Economic Review 14 (1): 97–106.

Siegel, Paul and Robert Hodge (1968), "A causal approach to the study of measurement error", in: Hubert Blalock and Ann Blalock, eds., Methodology in social research (McGraw-Hill, New York) pp. 28–59.

Siegfried, John J. and George H. Sweeney (1980), "Bias in economics education research from random and voluntary selection into experimental and control groups", American Economic Review 70 (2): 29–34.

Singer, Eleanor and Stanley Presser (1989), Survey research methods (University of Chicago Press, Chicago, IL).

Solon, Gary R. (1985), "Work incentive effects of taxing unemployment benefits", Econometrica 53 (2): 295–306.

Stafford, Frank (1986), "Forestalling the demise of empirical economics: the role of microdata in labor economics research", in: Orley Ashenfelter and Richard Layard, eds., Handbook of labor economics (North-Holland, Amsterdam).

Staiger, Douglas and James H. Stock (1997), "Instrumental variables regression with weak instruments", Econometrica 65 (3): 557–586.

Stigler, Stephen M. (1977), "Do robust estimators work with real data?" Annals of Statistics 5 (6): 1055–1098.

Stoker, Thomas M. (1986), "Aggregation, efficiency and cross-section regression", Econometrica 54 (1): 171–188.

Sudman, Seymour and Norman Bradburn (1991), Asking questions: a practical guide to survey design (Jossey-Bass Publishers. San Francisco, CA).

Taubman, Paul (1976), "Earnings, education, genetics and environment", Journal of Human Resources 11 (Fall), 447–461.

Taussig, Michael K. (1974), Those who served: report of the Twentieth Century Fund Task Force on policies towards veterans (The Twentieth Century Fund, New York).

Thurow, Lester C. (1983), Dangerous currents: the state of economics (Random House, New York).

Topel, Robert H. (1991), "Specific capital, mobility and wages: wages rise with job seniority", Journal of Political Economy 99 (1): 145–176.

Trochim, William K. (1984), Research design for program evaluation: the regression-discontinuity approach (Sage, Beverly Hills, CA).

Tufte, Edward R. (1992), The visual display of quantitative information (Graphics Press, Cheshire, CT).

Tukey, John W. (1977), Exploratory data analysis (Addison-Wesley Publishing Company, Reading, MA).

US Bureau of the Census (1992), Current population survey, March 1992. Technical documentation (Bureau of the Census, Washington, DC).

US Bureau of the Census (1996), Census of population and housing, 1990 United States: public use microdata sample: 5 percent sample. Third ICPSR release (US Department of Commerce,Washington, DC).

van der Klaauw, Wilbert (1996), "A regression-discontinuity evaluation of the effect of financial aid offers on college enrollment", Unpublished manuscript (Department of Economics, New York University).

Vroman, Wayne (1990), "Black men's relative earnings: are the gains illusory?" Industrial and Labor Relations Review 44 (1): 83–98.

Vroman, Wayne (1991), "The decline in unemployment insurance claims activity in the 1980s", UI occasional paper no. 91-2 (Employment and Training Administration, US Department of Labor).

Wald, A. (1940), "The fitting of straight lines if both variables are subject to error", Annals of Mathematical Statistics 11: 284–300.

Welch, Finis (1975), "Human capital theory: education, discrimination and life-cycles", American Economic Review 65: 63–73.

Welch, Finis (1977), "What have we learned from empirical studies of unemployment insurance?" Industrial and Labor Relations Review 30: 451–461.

Westergard-Nielsen, Niels (1989), "Empirical studies of the European labour market using microeconomic datasets: introduction", European Economic Review 33 (2/3): 389–394.

White, Halbert (1980), "Using least squares to approximate unknown regression functions", International Economic Review 21 (1): 149–170.

Willis, Robert J. and Sherwin Rosen (1979), "Education and self-selection", Journal of Political Economy 87 (5): S7–S36.

Yitzhaki, Shlomo (1996), "On using linear regressions in welfare economics", Journal of Business and Economic Statistics 14: 478–486.

Chapter 24

NEW DEVELOPMENTS IN ECONOMETRIC METHODS FOR LABOR MARKET ANALYSIS

ROBERT A. MOFFITT*

Johns Hopkins University

Contents

Abstract 1368
JEL codes 1368
1 What labor economists do 1369
2 Developments in qualitative, limited-dependent, and selection bias models 1374
2.1 Binary choice model 1374
2.2 Multinomial choice model 1382
2.3 Censored regression model (Tobit) 1387
2.4 Sample selection bias model 1389
3 Conclusions 1393
References 1394

* The author would like to thank Michael Keane and the participants of a conference at Princeton University in September 1997 for comments, and Julie Hudson for research assistance.

Handbook of Labor Economics, Volume 3, Edited by O. Ashenfelter and D. Card

Abstract

Econometric practice in labor economics has changed over the past 10 years as probit, logit, hazard methods, instrumental variables, and fixed effects models have grown in use and selection bias methods have declined in use. To a large degree these trends reflect an increasing preference for methods which are less restrictive, more robust, and freer in functional form than older methods, although not all trends are consistent with this view. The trends also reflect a tension between structural and reduced-form estimation that has not yet been resolved. A major point of the review is that this trend in labor economic practice has paralleled a trend in econometrics involving the use of flexible forms and semi-parametric and non-parametric methods but has not incorporated the lessons from that field.

JEL codes: J00; C1

Labor economists have long regarded their field as the most econometrically sophisticated of the various fields in microeconomics. Many of the major developments in microeconometrics in the 1970s – limited dependent variable and selection bias models, panel data models, hazard analysis, and so on – were often developed with labor economics applications in mind, although certainly not exclusively. Many of the econometric developments of the 1970s were stimulated by the new availability of data from household surveys, both cross-section and panel, which contained information on relatively large numbers of individuals. Many, if not most, of the types of issues that such datasets are best suited for study are found within the scope of labor economics. The development of computational hardware and software in the 1970s also grew rapidly, and this aided the development of many of the econometric methods which required somewhat more computational burdens than ordinarily least squares (OLS). Thus the confluence of econometric developments, data availability, and computational aids all contributed to the rapid advance in methodology in the 1970s in labor economics.

Volumes 1 and 2 of the *Handbook of Labor Economics* (Ashenfelter and Layard, 1986) appeared at a point subsequent to these major econometric developments. While there was no single chapter in that *Handbook* devoted to econometric methods, several of the chapters discussed econometric methods at least briefly as part of a particular topic area, such as those on labor supply (Pencavel, Killingsworth and Heckman), labor demand (Nickell), education (Willis), and hedonics (Rosen). The econometric issues discussed in these chapters were primarily the econometric developments of the 1970s just referred to. As is typically the case, debates in the profession concerning those methods can usually be detected only implicitly in print, for much of the debate over econometric methods is oral rather than written.

This chapter is concerned with econometric methods in labor economics and takes both a retrospective and prospective approach. Retrospectively, the use of econometric methods since 1986 is surveyed and discussed. Over the past decade there have been significant changes in the types of methods used in labor economics which, while mostly known to

applied labor economists who have practiced over this period, will be documented here for the first time. In addition, the discussion in this chapter will provide one author's view of why some of those trends have occurred. The chapter will then focus on a subset of the methods used by labor economists, namely, those for qualitative and limited dependent variables (probit, logit, Tobit) and those for selection bias models. The chapter will discuss prospects for the econometric developments in these areas in labor economics over the next 10 years. Perhaps the safest prediction is that methodological views will continue to evolve, for it does not appear to this author that econometric practice is now in equilibrium. Several new econometric developments, such as non-parametric and semi-parametric methods, are discussed and their prospects for increased use are discussed as well.

The results of the survey in the chapter reveal that labor economics is not quite as econometrically sophisticated today as it might be thought. The techniques which have seen the largest increase in usage in the top journals over the 1986–1996 period have been instrumental variables and fixed effects methods, both of which were essentially fully developed considerably prior to 1986. However, there has also been a slight increase in the number of more advanced probit and logit methods used, which may be testimony to lags in the application of econometric methods and/or to their introduction into software packages, for these methods were also developed prior to 1986. The number of applications of frontier econometric methods – simulation methods, non-parametric and semi-parametric methods, and the like – remains exceedingly small. Their use may, of course, be even smaller in other fields of applied economics, but those are not surveyed here.

More generally, the survey reveals that econometric practice in labor economics is shifting toward techniques that are, or at least can be argued to be, less restrictive and more robust than some of those used in the past. Identification of the parameters of econometric models has become a more important focus of attention than it has been historically. This is a trend which would appear to be occurring across many fields of economics, in other social science disciplines, and in fact in the field of statistics itself. It is safe to predict that this trend will continue.

1. What labor economists do

A survey of econometric methods used in labor economics should first determine what labor economists actually do. Table 1 shows the results of a survey of labor economics articles appearing in six highly-ranked general interest journals and two field journals in labor economics in 1985–1987 and then again in 1995–1997 to ascertain trends, using the labor economics classification scheme employed by the *Journal of Economic Literature*.

Interestingly, there has been a decline in the total number of labor economics papers published in those journals surveyed over the period. It is possible that this could be an artifact of the change in the *JEL* classification scheme between the periods but this, if anything, would work in the opposite direction because the scheme in the latter period was more inclusive (e.g., the economics of minorities and discrimination was moved into the

Table 1
Empirical and econometric work in labor economics, 1985–1997[a] (numbers of articles and methods)

	1985–1987	1995–1997
All articles	440	295
With empirical content	278	227
Without empirical content	162	68
Types of datasets used		
Cross-section	113	74
Panel	97	118
Repeated cross-section	40	46
Time series	18	26
Econometric methods used		
OLS	182	132
WLS	5	6
GLS	13	13
NLS	5	3
Probit	33	35
Ordered probit	3	5
Bivariate probit	1	5
Multinomial probit	1	3
Logit	29	23
Multinomial and conditional logit	5	17
Nested logit	1	4
Linear probability model	2	6
Simulation estimation	0	4
Tobit	21	25
Two-step selection bias	19	8
FIML selection bias	5	1
Other selection bias	6	10
IV and 2SLS	36	53
3SLS	4	0
Non-linear 2SLS	2	0
GMM	0	2
Fixed effects	27	56
Random effects	6	9
Hazard	21	25
Non-parametric	1	6
Semi-parametric	0	2

[a] Table surveys all articles appearing in 1985, 1986, and 1987 issues and 1995, 1996, and 1997 issues of the *American Economic Review*, *Econometrica*, *Journal of Human Resources*, *Journal of Labor Economics*, *Journal of Political Economy*, *Quarterly Journal of Economics*, *Review of Economic Studies*, and *Review of Economics and Statistics* which were listed under the labor economics headings in the classification codes of the *Journal of Economic Literature*.

labor economics category). The large majority of the decline, however, is revealed in Table 1 to have resulted from a decline in the number of theoretical articles published in the field, reflecting a genuine decline in theory in labor economics. This represents a reversal of the trend noted by Stafford (1986) of a marked rise in theory in labor economics from 1965 to the late 1970s and early 1980s.[1] Whether the decline in theory represents a return to the older research style in labor economics, with its institutional and non-theoretical orientation, remains to be seen.

There has also been a slight drop in the total number of articles published with empirical content – defined as having at least one table of non-artificial, non-simulated data – but this could be the result of fluctuations in the numbers of articles published each year, the particular years used in Table 1, and the particular journals chosen.

There have also been shifts over the decade in the types of datasets used in labor economics articles, also displayed in the table. The number using single cross-sections has drastically declined while the number using panel datasets has increased. Although there are few panel datasets available in the later period that were not available in the earlier one – the Survey of Income and Program Participation is an exception – the growing length of panels like the PSID and NLSY, together with the increased appreciation of panel econometric methods, is no doubt responsible for the marked increase in their use. There has also been a slight increase in the number of repeated cross-section datasets used – sometimes called pseudo-panels – primarily resulting from the growth in applications using the Current Population Survey (CPS). There has also been, perhaps surprisingly, an increase in the number of times articles have used time-series datasets, although the absolute number in both periods is far smaller than the numbers for the other dataset types.

The rest of the table shows the number of econometric methods of different types used in labor economics articles in the two periods. Individual articles can contribute more than one entry to the table if they used more than one technique, an issue to which I shall return momentarily.

While least squares is still the workhorse of empirical work in economics – the number of times it is used dominates the others in the table by an order of magnitude – there has nevertheless been a 30% decline in the number of times it has been used. The major sources of this substitution appear later in the table and will be seen to be growth in the use of instrumental variables and fixed effects methods.

Turning to qualitative dependent variable methods, the survey shows that the number of times probit methods of any type have been used has increased slightly, although most of the growth has been in the use of nonstandard variants such as ordered probit, bivariate probit, and multinomial probit. The continued popularity of probit is no doubt partly a result of the large number of labor economics dependent variables which are dichotomous – labor force participation, union status, migration, and educational categorizations, to name just a few – but also the degree to which probit has been incorporated into the major

[1] Manser (1999) detected a decline in theory in labor economics in a survey updating Stafford through 1993, however.

software packages used by applied economists. There has also been a slight increase in the use of logit-related methods, although here the shift away from simple binary logit to more advanced variants such as multinomial, conditional, and nested logit is more pronounced than for probit. While the more advanced logit techniques had been in existence for some time by the mid-1980s, they were still relatively new and their incorporation into econometric practice had not been completed. In addition, once again, these variants were not as widely incorporated into software packages as they are today, which no doubt is an additional contributor to the trend. An alternative hypothesis is that the types of topics which labor economists study have shifted toward types for which these techniques are most appropriate – that is, topics in which multiple discrete outcomes are the object of interest, such as occupational choice – but there is little evidence that there has been any significant shift of this kind.[2]

Table 1 also shows that there has been a slight increase in the use of the linear probability model in labor economics. While the number of times it is used is minuscule compared to what are clearly the more popular techniques of probit and logit, it has grown to a nontrivial number in the later period. This model will be discussed more in the next section and some reasons for this growth will be advanced.

Simulation methods, which are often used for the estimation of large-dimensional discrete choice models, have grown in use considerably over the period and will be discussed in the next section. The use of Tobit analysis has increased slightly but not as dramatically as some of the other methods, and it would be fair to characterize its use as relatively stable. It is a popular technique, used almost as frequently as probit.

Selection bias methods of all types have shown a marked decline over the period. This includes the two-step methods as well as full-information maximum likelihood methods. The reasons for this decline will also be discussed below. Moving in the opposite direction are methods using IV or two-stage least squares (2SLS), which have grown enormously. It will be argued below that these trends are related and are the result of a shift in econometric practice toward methods which require fewer distributional assumptions on unobservables, although it will also be argued that this is to some extent an oversimplification which ignores the less-parametric selection bias methods which have become available in recent years. The growth in IV and 2SLS is still quite remarkable given that those techniques have been widely used in economics for 30 years and had been developed long before that. While it could be argued that recent debates on the use of IV (e.g., Imbens and Angrist, 1994; Bound et al., 1995; Heckman, 1997; Staiger and Stock, 1997) have deepened the profession's understanding of the nature, interpretation, and limitations of IV and 2SLS, very few of the recently-discussed issues had not surfaced before in the econometric literature on these methods. Thus this trend, alone among those in the table, must be largely ascribed to a shift in the preferences of users rather than from the development of new econometric methods, to which the growth of other entries can be ascribed.

The growth in the use of panel datasets mentioned above is necessarily accompanied by

[2] Manser (Table A1, 1999), for example, finds no major shifts in the topics studied in labor economics articles over the past 10 years.

Table 2
Types of fixed effects models used in labor economics, 1985–1997[a] (numbers of times used)

	1985–1987	1995–1997
Individual	10	20
Family	3	11
Cohort	3	2
State	2	9
Geographic, non-state	2	5
Firm	2	3
Industry	1	0
School	0	2
Nationality	0	1
Other	2	1

[a] See Table 1.

a growth in the use of econometric methods for panels. The major growth has been in the use of fixed effects methods, whose use has doubled over the period. Table 2 shows, however, that not all of the growth in the use of this method can be traced to the increased use of panels. While individual fixed effects are indeed the modal category of use, and while models with fixed effects of that type have indeed grown more than those using any other type, the growth of models using family fixed effects and geographic fixed effects (state, city, country, etc.) has been equally dramatic. The resurgence of interest in sibling models, for example (Ashenfelter and Krueger, 1994; Behrman et al., 1994 to cite two examples among many) – models which have a long history in social science research – is part of this trend. The use of state fixed effects has been aided by the growth in the availability of more years of the CPS and of the growth in its sample size which makes estimation of state-specific intercepts more feasible.

The rest of Table 1 shows that the use of hazard methods – also called event-history, transition, or duration methods – has increased slightly, no doubt a combined result of the increase in data availability of panels and of the spread of knowledge and software incorporation of these techniques. Non-parametric and semi-parametric methods have grown significantly over the decade as well but still, in the later period, remain extremely limited in use. A major issue for the future is whether the use of these techniques will grow and become more common and, if so, at what level their use will plateau and what role they will come to play in the toolkit of techniques available to labor economists.

A close reading of Table 1 reveals that many more techniques have grown in usage than have declined. This reflects another trend in econometric practice in labor economics, which has been the growth in the number of multiple techniques used in the typical article. Table 3 shows the distribution of numbers of techniques used in different articles and shows this trend clearly, for the fraction of labor economics articles using only one method

Table 3
Number of different econometric methods used in labor economics articles, 1985–1997[a] (percent distribution)

	1985–1987	1995–1997
1 Method	59	40
2 Methods	26	33
3 Methods	11	16
4 Methods	3	9
5 Methods	1	1
6 Methods	0	0
Total	100	100

[a] See Table 1.

has dropped from almost 60% of all articles in the earlier period to 40% in the latter one. The offset is shown in uniform increased usage of two or more methods.

This last trend reflects a more basic underlying pattern affecting many of the other findings from the survey, of a movement toward the use of less parametric methods, more use of sensitivity testing and multiple methods to test for that sensitivity, and use of robust techniques that are less sensitive to assumptions. Much of applied thinking in labor economics practice today – and in the practice in many other fields inside economics and outside of it – is centered on these sets of issues. A safe prediction to make is that practice is not in steady state and that the trend in this direction will continue, although it would be hazardous to speculate on what exactly where it will be 10 years hence. Nevertheless, this will be the central theme of the rest of the paper, which will discuss trends and developments in qualitative and limited dependent variable models and selection bias models.

2. Developments in qualitative, limited-dependent, and selection bias models

The remainder of the chapter focuses on developments in several of the techniques which have changed in usage and in which thinking has developed considerably over the last 10 years. These are models for qualitative and limited-dependent variables, and selection bias models. The binary choice model will be discussed first and most exhaustively because, despite its simplicity, many of the developments in the other models can be seen most easily and simply when the outcome is dichotomous.

2.1. Binary choice model

2.1.1. Basic considerations

The most popular binary choice framework in labor economics is the probit model, which can be written (suppressing individual-observation subscripts):

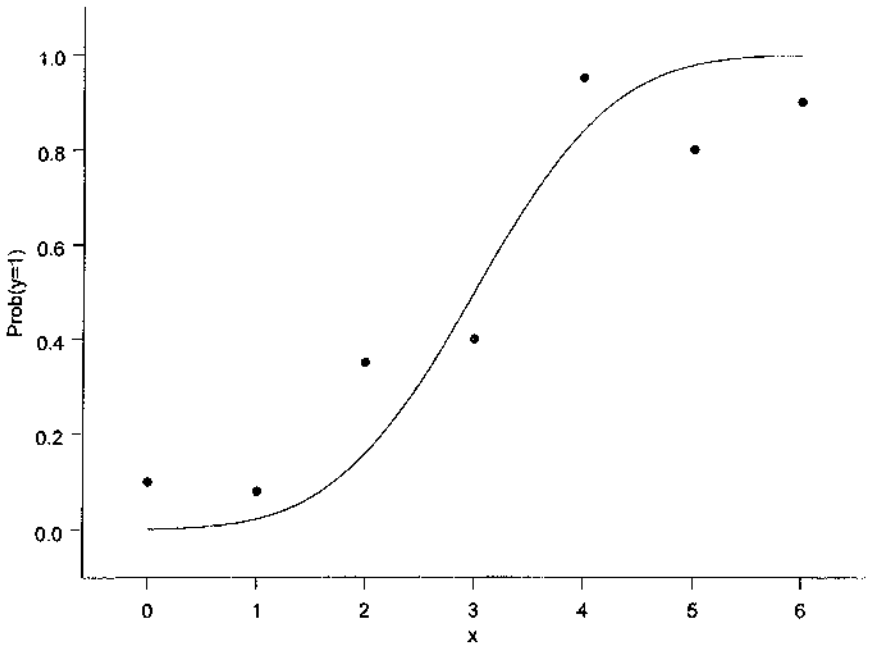

Fig. 1. Probit probability curve with single x with positive coefficient. Dots represent sample means of y.

$$y^* = X\beta + \varepsilon, \tag{1}$$

$$y = 1 \text{ if } y^* \geq 0, \qquad y = 0 \text{ if } y^* < 0, \tag{2}$$

$$\varepsilon \approx N(0, 1), \tag{3}$$

where X is a row vector of variables an observation and β is a column vector of coefficients. The usual normalizations are imposed here – the variance of the error is normalized to 1 and the cutoff point is normalized to zero implying, respectively, that the coefficient estimates are ratios to standard deviations and that the estimated intercept absorbs the true (constant) cutoff. The model implies that

$$\text{Prob}(y = 1 \mid X) = \text{E}(y \mid X) = F(X\beta), \tag{4}$$

where F is the cumulative normal distribution function and is shown as the familiar S-shaped (sigmoid) curve in Fig. 1.

The probit model has a long history – the treatise by Finney (1947) claims that it originated in 1860 – and has been primarily used in the study of bioassay. In the social sciences it has also been extremely popular as a method of estimation of models with dummy dependent variables. While its use in bioassay is based on the theoretical notion

that the susceptibility of a group of experimentally-treated subjects to a stimulus has a distribution of responses, the technique is mainly used simply as a curve-fitting method ("...a convenient way...of fitting a function" (Berkson, 1951, p. 327)). Its appeal in economics is based instead on developments in the psychological literature which conceived of human subjects as having a scalar index representing an unobserved propensity, or utility, of an alternative; that differences in the index could be scaled; and that there exists a distribution of those propensities in the population (Thurstone, 1927). This led in turn to the familiar random utility model (Quandt, 1956; McFadden, 1974) as a justification for probit and other binary choice models. More generally, one of its major attractions to economists arises because choices of workers and firms are generally regarded as based on optimizing behavior resulting from utility or profit maximization, and that y^* can be thought of as the difference between the utilities or profits of alternative choices.

The other popular model is the logit model (Berkson, 1944), in which

$$\text{Prob}(y = 1 \mid X) = \frac{e^{X\beta}}{1 + e^{X\beta}}. \tag{5}$$

This model can also be generated by a latent index model such as (1) where the values of each alternative have error terms which are distributed independent extreme value and where y^* again represents the difference in those values and therefore fits just as well into the random utility model as probit (McFadden, 1974). There is an oral tradition in applied microeconomics that probit and logit estimates are almost always close to one another, which is based on the result that the c.d.f.'s of the two normalized distributions are quite close to one another except in the tails (see, e.g., Domenich and McFadden, 1947, p. 58). The oral tradition that the coefficient estimates are not much affected is, however, a somewhat different result and depends on the configuration of the data and the distribution of y and X. Nevertheless, it is based on empirical findings from a large number of applications where both probit and logit have been estimated and hence has a grounding in typical applications.[3]

The linear probability model (LPM) – that is, the model which posits a linear relationship $y = X\gamma + e$ and interprets γ as the effect of a unit change in X on the probability that $y = 1$ – has been fairly uniformly criticized in econometrics textbooks, both older and current. The most cited discussion among the early econometric textbooks is that in Goldberger (1964), who noted that the LPM could produce fitted values of y outside the zero–one range and that e was necessarily heteroskedastic. This criticism appears in the textbook by Theil (1981) and in current textbooks as well. From the current vantage point neither of these disadvantages seems fatal to the LPM because fitted values of y close to the mean of the data are unlikely to lie outside the zero–one range, and thus do not cause a

[3] At one period in the early 1970s it was humorously thought that logit seemed to be preferred by West Coast economists and probit, by East Coast economists. Today the choice between the two is generally thought to be entirely a matter of taste. It was not always so, for in the early development the difference was seriously debated (e.g., Berkson, 1951).

problem if prediction or inference only close to the mean is desired; and because there are a variety of methods now available for the correction of heteroskedasticity.[4]

The reason the LPM has seen a slight return to use is not that anyone believes it should be taken to be the true model – in which case there are many statistical objections to it – but that it can be thought of merely as a linear approximation to the true model, and as a reduced form.

According to the random utility model,

$$y = \mathrm{E}(y \mid X) + u = F(X\beta) + u, \tag{6}$$

where u is a mean-zero, albeit heteroskedastic and bounded, error term. The LPM can be thought of merely as a linear approximation to the non-linear function $F(X\beta)$ and its coefficient estimates are unambiguously and correctly interpretable as such. The coefficients estimated in the LPM are direct estimates of $\partial \mathrm{E}(y \mid X)/\partial X$ and hence do not estimate β but rather some combination of β and F. If one is not particularly interested in estimating the index function coefficient (i.e., β) itself, but only in the net, or reduced-form, effect of a change in X on the probability that $y = 1$, the LPM is arguably an acceptable place to start.[5]

The case for the LPM in this respect must therefore rest partly on whether the question being asked by the analyst requires an estimate of the underlying index coefficient β or just the reduced-form effect of X on $\mathrm{Prob}(y = 1 \mid X)$. The random utility theories which underlie so much of the work in this area make β the object of interest and it is therefore natural to seek estimates of that parameter. A contrary argument can be made that there are not many labor economics hypotheses that depend on the magnitude of β and, if it can be assured that the signs of β and of $\partial \mathrm{E}(y \mid X)/\partial X$ are the same, hypotheses on the sign of the former can be tested from the estimated sign of the latter. If the contrary argument is accepted, the remaining question is whether significant non-linearities are present in the data that are missed by a linear approximation. As with OLS in general, the relevant information in the data used for the coefficient estimation is the set of means of y at each value of x (Fig. 1 includes some scatter points of those means which, according to the model, are generated by $F(X\beta)$ plus a mean of u at each x). The non-linearity of the curve varies over the range of x and the second derivative of the function, which is one measure of non-linearity, reaches an maximum in absolute value at two points on the function; the linear approximation is likely to be the worst in those regions and it is likely to be the best if the data are tightly clustered in either of the two tails or around the 0.50 probability point, where the curve has the lowest curvature. Data configurations which

[4] Fitted values outside the zero–one range do nevertheless cause problems of interpretation of the R-squared and also problems with heteroskedastic fixups such as weighted least squares. Another criticism of the LPM is that it is inconsistent with the random utility model. See Heckman and Snyder (1996) for a discussion of this point.

[5] Amemiya (1981, p. 1487) recommends the LPM in the early stages of an analysis as a convenient way to summarize the data and to gain quick approximate estimates of relationships, but believes that a more formal binary choice model should be used for the final analysis. This is no doubt the way it is used by most applied economists.

have more dispersed sample probabilities over the unit interval are likely to be poorer approximations because more non-linearities are present in that case.[6]

If non-linearities are an issue, a logical alternative procedure is non-parametric (NP) regression (Härdle and Linton, 1994), for NP regression can capture arbitrary non-linearities in a fitted curve. NP regression of y on X in one of its standard forms (e.g., kernel regression or series regression) can be applied without modification even if y is binary, for the method merely fits the means of y in the data at each x, $E(y \mid x)$, to a non-linear curve; that the means of y are fractions is irrelevant. A less drastic solution would be to introduce polynomials, splines, interactions, and other forms of non-linearities in X but within the framework of the basic linear-in-parameters model so that OLS could still be applied. Unfortunately, this approach is not feasible if X is high-dimensional and hence is not a practical alternative.

However, NP regression is still not widely used in labor economics or in other applied areas of economics despite its attractiveness as a method of capturing non-linearities. There is no consensus on why it has not been incorporated in econometric practice more than it has been. NP regression has now been widely known and heavily researched among econometricians (not to mention the wider community of statisticians) for at least 10 years. One simple reason may be that the lags in the incorporation of new econometric techniques into practice is still quite long. A related reason may be that new techniques may have to be incorporated into software packages and that NP regression has not, to date, been incorporated into any of the major packages used by economists.[7] A more problematic reason for the lack of use of NP regression is that estimation can still be somewhat computationally burdensome, especially if there are large numbers of regressors; that large sample sizes are often needed for estimates without major bias, and that rates of convergence might be low; and that the choice of bandwidth has an impact on the estimates but the preferred method of choosing bandwidth has not, to date, been sufficiently standardized through rules of thumb and guides to practice. Unfortunately, the theory of bandwidth selection implies only that it should go to zero asymptotically, as well as satisfying a few other general criteria (Manski, 1991, pp. 43–44), which is not enough of a guide for the practitioner community. A body of practical experience must be built up and widely-agreed upon rules of thumb will have to be adopted instead. It is possible that this will occur in the future and that more use of NP regression will be consequently seen.[8]

[6] See Aldrich and Nelson (1984) for further discussion of using the LPM as a linear approximation to the true curve.

[7] Some packages have non-parametric density estimation, which can be used to estimate NP regressions. Also, LIMDEP permits the estimation of the maximum score model (see below) and some related models.

[8] See the recent article by Blundell and Duncan (1998) for a start in this direction. It should also be noted that some econometricians argue that the semi-parametric models to be discussed momentarily – which permit some forms of non-linearity but still retain some parametric structure – should be considered the viable alternative to NP regression, given the problems with the latter, even for exploring non-linearities and for estimating the reduced-form function $E(y \mid x)$. How to evaluate semi-parametric versus non-parametric models in that light has not been articulated, however. This argument is to be sharply distinguished from the purpose of semi-parametric models discussed in the next section, which is to identify β.

2.1.2. Further issues: heteroskedasticity, identification, and policy experiments

Assuming that one is interested in β itself, and not just in the reduced-form effect of X on y, probit or logit remain the main alternatives currently used by the analyst. Unfortunately, the recent literature on these techniques highlights important restrictions they impose that are often only implicit, and which raise some discomfiting issues that are not easily resolved. One restriction is that the distribution of the unobservables is normal or logistic, and violation of this restriction results in inconsistent estimates of β. Interestingly, however, the literature to date indicates, at least for the binary choice case, that reasonably minor deviations from the normal or logistic – for example, deviations that remain in the class of unimodal distributions – do not much affect the estimates (Manski and Thompson, 1986; Horowitz, 1993b). To obtain a major change in the magnitude of estimated β from this source requires that the error term be distributed very differently, e.g., to be bimodal rather than unimodal (Horowitz, 1993b).

What appears to be a potentially more serious problem is heteroskedasticity, which also (unlike the linear model) results in inconsistent probit and logit coefficients. One approach to the heteroskedasticity problem is to allow for some parametric form of heteroskedasticity, e.g., $\mathrm{Var}(e \mid X) = (X\delta)^2$, and to build this into the likelihood function for probit (or logit). An alternative is to conduct specification tests for such a form of heteroskedasticity or for a more general type; there are a wide variety of such tests available (see Pagan and Vella, 1989; Maddala, 1995, for reviews). Nevertheless, the econometric literature in this area has revealed that a fundamental identification problem lurks if the functional forms are relaxed beyond a certain point. The expected value of y in Eq. (4) relies on two functional form assumptions: (*i*) that F is the normal c.d.f. and that the parameters of the normal distribution do not depend on X, and (ii) that the index is linear in β, as in $X\beta$. These assumptions are generally ignored in linear-model estimation not because they are thought to be true but because OLS coefficients are consistent if they are not (in the case of heteroskedasticity) or are a matter of convenience which could easily be relaxed (in the case of linearity). On the contrary, heteroskedasticity in labor applications could be pervasive, at least to some degree, and non-linearities are often uncovered in studies of the effect of education on earnings, wages on labor supply, and other traditional topics in labor economics.[9]

In the absence of either of these two functional form assumptions, the means of y at each X are given by

$$\mathrm{E}(y \mid X) = F[h(X, \beta); X], \tag{7}$$

where $F(u; X)$ is a proper c.d.f. whose parameters depend on X (i.e., heteroskedasticity is present) and $h(X,\beta)$ is an unknown function which is the main object of interest. It is clear

[9] There is no consensus on whether existing evidence supports serious concern with heteroskedasticity in these models. For an application showing specification tests which reveal no evidence of heteroskedasticity, see Melenberg and Van Soest (1996); for a study reporting specification tests which fail to reject sizable biases from heteroskedasticity, see Horowitz (1993a).

from (7) that one could never hope to disentangle F from h using only the pairs of y-means and X in the data. In Fig. 1, for example, it is impossible to know whether the rough increase in the mean of y as x increases is a result of a true change in the latent index h, or merely a result of a change in the distribution of the error term as x changes. One could fall back on the reduced-form approach, giving up on separating h from F and merely regressing y on X with either OLS or NP regression, but that does not solve the problem; it is still the case that the estimated effect of X on y may merely be picking up heteroskedasticity.

A sizable econometric literature has been built up around this and related issues (see Manski, 1988; Horowitz, 1993b; Powell, 1994, for reviews). One early treatment of the identification problem is addressed by Manski (1988), who termed (7) a "structural" model because it contains an unobserved latent index, as opposed to the reduced-form model $E(y \mid X) = G(X)$ which combines F and h. Estimation of structural models, in general, requires identification restrictions and this model is no exception, and the Manski paper as well as the later reviews discuss a variety of different restrictions that can be imposed on F and/or h to be able to identify β. One can, for example, take a position on the functional form of the index, e.g., $h(X, \beta) = X\beta$ and leave the form of F to be dictated by the data. This is sufficient for identification in what are known as "single index" models, although the form of any heteroskedasticity in those models must be further restricted. The maximum score estimator (Manski, 1975) is also in this class and allows arbitrary heteroskedasticity but at the price of assuming that the median of u is independent of X, and using only the sign of $X\beta$ to determine where y is 1 or 0, thereby using minimum information in the data. The maximum score estimator is one of the few that has been packaged (LIMDEP). Alternatively, one can assume normality (or some other distribution) for F and then let $h(X,\beta)$ be free to be determined by the data which, together with some other restrictions, is also sufficient for identification.[10] None of these alternative approaches is particularly attractive because they convert linearity in $X\beta$ and normality, respectively, from assumptions of convenience to assumptions necessary for identification which are not relaxable (within their class). Assuming linearity of $X\beta$, for example, is tantamount to ascribing all deviations from linearity to F. In addition, the estimation techniques which have been devised for these models have not, as of this writing, been standardized in a way that has made it easy for practitioners to estimate them easily, even as a side test to the robustness of probit or logit. Neither has much empirical experience has been built up upon which standardization could be based. For some techniques, some estimation has been conducted (e.g., for the maximum score estimator) but the empirical experience has not been particularly encouraging. The convergence rates of some of the estimators is also quite slow.

Perhaps the strongest restriction that can be imposed on (7) is simply to assume independence of u and X and therefore to assume that heteroskedasticity is not present. Under

[10] Another branch of this literature which is evolving is one which permits the estimation of a type of function $h(X,\beta)$ which is not fully parametric, as in $X\beta$, but partially parametric such as, for example, representing h as a sum of separable but unknown functions of different variables in X. This permits non-linearities in the effect of X to be estimated without going to the complete non-parametric approach.

that assumption there is still an identification problem because separating F from h requires some restrictions. Matzkin (1992) has provided the most well-known exposition of identification when both F and h are unknown but independence is maintained. Computation is problematic, however, and additional restrictions are required. An alternative is to retreat further and assume either a functional form for F (e.g., normality) or a functional form for h, either of which would simplify matters and make identification easier. Estimation in these cases would still require some non-parametric or related estimation method. One of the few empirical applications of this type of semi-parametric model is that of Newey et al. (1990), who estimated a labor-force-participation equation for women with probit and with two semi-parametric methods which assumed linearity of $X\beta$ and independence of X from the error term, but was non-parametric on the form of F. The authors found that the semi-parametric coefficient estimates were statistically no different than those from the conventional probit model. Still another approach is to assume a functional form for F that is more general and more flexible than the normal or the logistic and thereby move some distance toward the semi-parametric approach. One of the easiest approaches of this type is to assume that the distribution of the error term is a weighted sum of independent normals rather than a single normal (e.g., Geweke and Keane, 1999); this allows the probit model to be a special case but also allows F to take on a wider variety of shapes, for weighted sums of normals can capture many different types of distributions.

The independence assumption is the most attractive one for applications where a policy experiment is the main object of interest, for a genuine policy experiment is by definition one in which a particular X changes over time, or varies cross-sectionally, in a way that is independent of the underlying distributions of unobservables for different values of X. The heteroskedasticity problem, if interpreted as arising from preference heterogeneity correlated with X, for example, is essentially a problem of non-experimental methodology; to avoid it requires changing values of X while holding the types of populations being made subject to each value of X unchanged. This is achieved in a true randomized trial because both experimental and control group error distributions are, on average, the same aside from the effect of the treatment.[11] Indeed, the early literature on the development of probit analysis which took place in bioassay was explicitly experimental in focus, and hence heteroskedasticity was rarely explored.[12]

Where this leaves the practitioner is still somewhat in limbo. If estimation of β itself is desired, then if arbitrary assumptions on distributional form and linearity of the index are to be avoided, difficult and yet-to-be-standardized, or packaged, techniques are still required. There is some evidence that the normality assumption is not damaging as long as the true distribution is unimodal which, if this is maintained, would allow the investigator to simply introduce parametric non-linearities into the index function (still assuming

[11] It is assumed here that (1) is still the true causal model; that is, that there is no treatment effect on any part of the distribution other than the mean. If there is, the model must be modified and different parameters of interest must be introduced.

[12] Finney (1947) urged his readers to read the classic experimental text of R.A. Fisher (1925) before using probit analysis, for example.

independence as well). Moreover, many applied labor economists take the view that it has still not been demonstrated that probit and logit are not, in fact, very robust methods which will generally give approximately correct answers; there is no widespread evidence as yet that contravenes this view. Thus a defensible position at the present time is still to use one of the conventional techniques. On the other hand, if β itself is not of interest, either the LPM or, preferably, non-parametric regression is probably the current or possibly soon-to-be-current standard. These standards could, and probably will, change over the next several years.

2.2. *Multinomial choice model*

In the multinomial choice model, outcomes consist of multiple discrete categories rather than only two. For example, in a study of occupational choice there will be as many choices as there are occupations. A distinction worth making at the start is that between sets of choices which are mutually exclusive and sets which are not. The occupational choice example is clearly one with mutually exclusive choice, but a different case is the hedonic model in which the individual chooses jobs which have multiple discrete characteristics (pension/no-pension, health insurance/no health insurance, etc.). The latter can be converted into a mutually exclusive set of outcomes by crossing all the individual discrete outcomes with each other and, in so doing, generating a mutually set of combinations of job characteristics, but often this is not desired. This discussion will concentrate on the more common mutually-exclusive case. The non-mutually exclusive case should be thought of more in the class of multiple equation models like the seemingly unrelated regression model.

In the popular multinomial-conditional logit model, individual i must choose from among $j = 1, \ldots, J$ alternatives. Define y_{ij} as a binary variable equal to 1 if the individual chooses j and 0 if not. Then the model proposes that the probability that individual i chooses alternative j is[13]

$$\text{Prob}(y_{ij} = 1 \mid X_i, Z_j) = \frac{\exp(X_i\beta_j + Z_j\delta)}{\sum_{k=1}^{J} \exp(X_i\beta_k + Z_k\delta)}. \tag{8}$$

In Eq. (8), a distinction is made for clarity between variables that vary across individuals, X_i (race, sex, etc.), and variables that vary across alternatives, Z_j (e.g., characteristics of an occupation). In order for the X_i to have a sensible effect on the probability that individual i chooses j, it is necessary that the coefficient on X_i (β_j) vary across alternatives; otherwise, multiplying the top and bottom of (8) by $\exp(X_i\beta)$ would eliminate it from the model and it

[13] Individual subscripts i are added in this section but in no other in the paper. They are shown here because the distinction between X and Z, which is important for identification of the multinomial choice model, is less clear without the individual subscripts.

would not be estimable. Variables which vary across alternatives, like Z_j, on the other hand, can have a constant coefficient (δ).

As in the binary choice model, part of the appeal of the logit model for economists is that it can be derived from a random utility model in which utility maximization is assumed. As shown by McFadden, (8) is the probability that results from a choice problem in which an individual obtains utility from each alternative equal to

$$V_{ij} = X_i\beta_j + Z_j\delta + \varepsilon_{ij} \tag{9}$$

and in which the alternative j with the maximum value of V_{ij} is chosen. The form in (8) requires that the errors ε_{ij} be independently and identically distributed extreme value across alternatives for each individual i. The independence assumption generates the well-known property of independence of irrelevant alternatives (IIA) in the model, namely, that the ratio of the probabilities of choosing any two alternatives is independent of the parameters and the variables (Z_j) for all other alternatives, as can be seen by taking the ratio of two probabilities of the form of (8) for two alternatives j and j'.

The IIA problem and consideration of alternatives that do not require it has dominated the discussion of multinomial choice in applied econometrics in the last 10 years (and, in fact, for some period prior to that). The typical econometric discussion of IIA states that the assumption is violated if some of the alternatives are close substitutes, as would be the case if the individual were choosing between red and blue buses (an example due to McFadden). This is a little misleading because the IIA problem should be thought of more generally as a problem of correlated error terms or, perhaps easier to relate to, as a selection bias problem that arises whenever selecting subsamples of a population leads to inconsistencies in parameter estimates. The latter interpretation makes use of the implication of the IIA assumption that the model can be consistently estimated simply by using for estimation only the individuals in the sample who select one of two of the alternatives, say j and j', and by analyzing their relative choice with binary logit. As should be familiar from the general principles of sample selection bias, estimation on such a subpopulation may yield biased and inconsistent parameter estimates if the subpopulation that chooses only j or j' is systematically different from the rest of the population. The subsample choosing j or j' is a self-selected sample and their relative choices between the two alternatives are likely to be different than the choices that the rest of the population might make. Hence estimates based on the subpopulation will not yield parameters β_j and δ which apply to the total population. The IIA assumption presumes this not to be true; that the rest of the population would make the same relative choices between j and j'.[14]

The underlying issue is how to estimate choices when the ε_{ij} are correlated across j, as one would expect them to be in almost any occupational choice, job choice, or other labor

[14] Indeed, one of the tests for the IIA assumption (Hausman and McFadden, 1984) is based exactly on this formulation, for the test involves comparing the coefficients from the logit estimation on the full sample with the estimates obtained on subsamples like that choosing only two of the alternatives. Under the IIA assumption, the two methods should yield the same coefficient estimates. See Maddala (1995) for a review of this and other specification tests for the IIA assumption.

application where the individuals making the choice have unobserved preferences, or unobserved variables more generally, which are correlated across those alternatives (it would be surprising, in general, if they were not). One approach is to give up on the estimation of the structural model and to seek a reduced form which does not impose any independence on the errors across alternatives. A linearized reduced form might lead to a counterpart to the LPM in the binary choice model, for example. This option is feasible is there are no alternative-specific regressors (Z_j) but may not be if the number of Z_j is large. According to the choice model, an alternative j is chosen if the utility differences between it and other j' are all the same sign, i.e.,

$$V_{ij} - V_{ij'} = X_i(\beta_j - \beta_{j'}) + (Z_j - Z_{j'})\delta + u_{ijj'}, \tag{10}$$

$$u_{ijj'} = \varepsilon_{ij} - \varepsilon_{ij'}, \tag{11}$$

$$\text{Choose } j \text{ iff } V_{ij} - V_{ij'} \geq 0 \qquad \forall j'. \tag{12}$$

Therefore the reduced-form probability of choosing alternative j is

$$\begin{aligned} \mathrm{E}(y_{ij} \mid X_i, Z_1, \ldots, Z_J) &= \mathrm{Prob}(y_{ij} = 1 \mid X_i, Z_1, \ldots, Z_J) \\ &= \mathrm{Prob}(u_{ij1} > W_{ij1}, \ldots, u_{ijj-1} > W_{ijj-1}, u_{ijj+1} > W_{ijj+1}, \ldots, u_{ijJ} > W_{ijJ}) \\ &= g(X_i, Z_1, \ldots, Z_J), \end{aligned} \tag{13}$$

where

$$W_{ijj'} = -X_i(\beta_j - \beta_{j'}) - (Z_j - Z_{j'})\delta. \tag{14}$$

Thus the reduced form for the choice of j must contain as arguments not only X_i and Z_j but also the $Z_{j'}$ for all other j'. That is, the probability of choosing j is a function of all characteristics of all alternatives. If the application involves a large number of alternatives, or if there are very many variables Z_j for each alternative, this yields a model with an impractical number of independent variables. Moreover, the large number of coefficients obtained from estimating such long regressions separately for every alternative would be an inefficient method of estimating the model compared to any alternative that recognizes that there is a much smaller number of underlying structural coefficients which are determining all the reduced form coefficients.

Nevertheless, if there are no Z variables in the application or if the number of Z variables or alternatives is small, the reduced-form approach is quite feasible. A LPM which projects y_{ij} onto all X_i and all Z_j, or an NP regression which does the same but better captures the non-linearities involved, are interpretable approaches.[15] There is also an

[15] Independence of the distribution of the errors from X is assumed throughout. As in the binary choice model, for example, heteroskedasticity would adversely affect any of these reduced-form approaches as well as structural approaches.

approach called "universal" logit which assumes (13) to have a logit form in which X_i and all Z_j enter the model.[16] In this case the logistic assumption is arbitrary and does not follow from the underlying error structure of the ε_{ij}, but is simply a way of capturing nonlinearities and keeping the dependent variable within the unit interval. In all these approaches, because reduced-forms are estimated, the random-utility interpretation is lost and none of the estimated coefficients can be directly related to those in (9). Also, as in the binary choice model, the value of these approaches depends on whether direct knowledge of the parameters of (9) is of interest rather than reduced form estimates of $\partial \mathrm{E}(y \mid X, Z)/\partial X$ and $\partial \mathrm{E}(y \mid X, Z)/\partial Z$.

Other approaches to the IIA problem retain the object of interest as estimating the parameters of (9) and hence are structural to some degree. The nested multinomial and generalized extreme value (GEV) models (McFadden, 1981), which have increased somewhat in popularity as noted in Section I, are of this type. In these models it is necessary to be able to assign the alternatives to a tree, or sequential, structure in which some of the alternatives are chosen after (in a temporal sense) or independently of (in a more general sense) some of the other alternatives. One application of this approach is to assume that a woman first chooses whether to work and only then whether to work part-time or full-time; another is that an unmarried woman with a child first decides whether to marry and only then, if she does not, whether she will go onto welfare. The nested logit and GEV models permit a degree of correlation between the error terms of the equations for the value of the lower-level alternatives, while maintaining independence from the upper levels. Unfortunately, as the two examples just given illustrate, the behavioral assumptions involved are strong and may be untenable. Most women undoubtedly jointly choose whether to work or not, and whether to work part-time or full-time; the choices are not sequential or separable. Nevertheless, these models have a role to play at least as a specification test for the fully independent model and are often worth estimating (the nested logit model is available in software packages)

An alternative approach that has undergone additional discussion in the last several years is multinomial probit with correlated errors. In this model, (9) is assumed to be the correct specification of utility for each alternative but the ε_{ij} across j are assumed to be distributed multivariate normal with a relatively full covariance structure (i.e., with nonzero correlations between the ε_{ij}). The probability of choosing each alternative is again in the form of (13) but now the aim is to actually evaluate that probability under the assumption that the underlying errors are multivariate normal. The problem in this case is entirely a computational, or numerical, one, for evaluation of high-order multivariate normal probabilities was long considered computationally infeasible even with modern hardware. However, Lerman and Manski (1981) showed that such probabilities could nevertheless be numerically computed by means of Monte Carlo simulation methods in which random draws from a multivariate normal distribution are repeatedly taken to form an estimate of the probability in question. Later work by McFadden (1989) and Pakes and Pollard (1989)

[16] See Amemiya (1985, p. 307) for a discussion of this model.

demonstrated the consistency and other properties of this and related estimators. A variety of alternatives have developed – methods of moment and maximum likelihood simulation methods – and a sizable literature has grown up around them. Several surveys are now available which outline the various approaches to estimation that have been developed (Hajivassiliou, 1993; Keane, 1993; Hajivassiliou and Ruud, 1994; Stern, 1997) and the methods have been extended to panel data (Keane, 1994).

To date these techniques have not been utilized to the extent that their potential would allow.[17] Multinomial logit is still by far the norm in estimation of multinomial choice models. A major reason for this lack of use is probably entirely practical, namely, that simulation methods have not been incorporated into software packages or standardized sufficiently to allow their routine use by applied economists. While writing a program to conduct the necessary numerical computations is in principle not difficult, it is sufficiently time-consuming as to be beyond the time capacities for most applied work. When and if the software firms incorporate these simulation techniques into their products will probably largely determine when and whether these techniques will spread in use.

A second and possibly more serious problem that has received some attention in the econometric literature is the identification of the multinomial probit and other models with correlated errors and, in particular, the identification of the across-alternative correlation coefficients that are at the heart of the contribution of multinomial probit over multinomial logit.

In the linear model, cross-equation correlation coefficients and covariances can usually be estimated from the sample covariance of residuals across equations, but in this case no similar approach would be feasible because the covariance of y_{ij} and $y_{ij'}$ is identifiably zero for all pairs – an individual is observed to choose only one alternative by definition. Therefore it is difficult to see how one could ever estimate a correlation in unobserved tastes (for example) between alternatives. The cross-equation correlations must instead therefore be identified from the conditional mean function g for each choice shown in (13), which relate the choice of each alternative to the X_i and the $Z_{j'}$ for all alternatives j'. The functional form in which the $Z_{j'}$ enter the g functions will differ depending on whether the $\varepsilon_{ij'}$ are or are not independent across all j', and it is this difference that must furnish identification. In the case where there are no Z_j at all, and hence each g function in (13) is simply a non-linear function of X_i, it is clear that no correlation coefficients could be identified in a completely distribution-free specification of the ε_{ij} and that all identification would come only from the non-linearities inherent in the multivariate normal distribution. Identification seems more possible if Z_j exist because the relation between y_{ij} and the $Z_{j'}$ for other alternatives should provide some information on the correlation. One consequence of this problem is that estimates of the multinomial probit model appear to be quite sensitive to the existence, and choice, of alternative-specific variables, as demonstrated by Keane (1992) and Geweke et al. (1994). This problem has not been completely worked-out in the econometric literature.

[17] For two examples of labor economics applications to date, see Berkovec and Stern (1991) and Keane and Moffitt (1998).

The non-parametric and semi-parametric literature has also not yet addressed multinomial models in depth. An extension of the maximum score model mentioned previously to the multinomial case has been proposed but has been little used in practice (Manski, 1975). The identification and estimation problems that arise when the normality or homoskedasticity assumptions are dropped have also not been extended to the multinomial model yet as well (see Horowitz, 1993b, and Powell, 1994, for references). Development of practice in labor economics in this direction must therefore await more progress in the econometric literature.

2.3. Censored regression model (Tobit)

The survey in Section 1 revealed that the Tobit model has retained its popularity in labor economics for the last decade. The model in its simplest form can be stated as

$$y^* = X\beta + \varepsilon, \tag{15}$$

$$y = y^* \text{ if } y^* \geq 0, \qquad y = 0 \text{ if } y^* < 0, \tag{16}$$

with $\varepsilon \sim N(0, \sigma^2)$. One of the most common uses of the Tobit model is in the analysis of the labor supply of married women, of whom a significant fraction do not work at any given point in time.

The general popularity of the model in labor economics, as well as in economics in general, is, as in the binary choice model, traceable to its easy interpretation in terms of individual and firm choice, where y^* represents either the demand or supply of a good, which will equal zero at a theoretically well-defined corner solution, or y^* is simply some other continuous choice which includes the option of not engaging in the consumption or activity at all.[18]

Despite the popularity of the Tobit model, the econometric literature on the model (there usually called the censored regression model) as it has developed over the last decade has revealed its fragility in the face of its assumptions. One source of fragility is the assumption of homoskedasticity which, as in the probit model, is necessary for consistent estimation (Hurd, 1979; Arabmazar and Schmidt, 1982; Brown and Moffitt, 1983). Monte Carlo evidence suggests that the asymptotic bias can be quite large. A second source of difficulty is the distributional assumption of normality for the unobservables, a problem examined explicitly by Arabmazar and Schmidt (1982) and Goldberger (1983). These papers show that the use of different distributions than the normal can yield quite different coefficients.

While both of these problems were present in the probit and logit models, they are potentially more severe in Tobit. In the two binary choice models, the dependent variable has a limited range from zero to one and there is considerable evidence that the S-shaped curves followed by a moderately wide range of different distributions are not far different, at least in the implications for the coefficient vector on a latent index. In the Tobit model,

[18] The Tobit model is due to Tobin (1958). See the September–October 1997 issue of the *Journal of Applied Econometrics* for a series of papers replicating and extending the demand function studied by Tobin.

on the other hand, the availability of continuous data on y leads, paradoxically perhaps (given that more information in the data must be regarded as better than less information), to a greater susceptibility to misspecification. The requirement that that conditional distribution of y, for those with positive y, be truncated normal and the same across individuals is a strong assumption that is commonly violated in many applications.

Although the labor economics literature has not yet absorbed the lessons of these results, the potential failure of the normality assumption in Tobit does have a counterpart in the labor supply literature in one area, which is the long-standing recognition that the distribution of hours of work per week and over longer periods as well, is highly clustered and decidedly non-normal (e.g., Pencavel, 1986); and that the determinants of the decision to work may be different than those for the choice of hours conditional on working. Hours per week are strongly clustered around 35–40, for example, and attempts to fit the conditional hours distribution to a truncated normal results in a poor fit to the fraction of those working zero hours; it is difficult to use the normal distribution to fit both. Fixed costs of work, which have been incorporated into econometric models of labor supply for some time (Hausman, 1980; Cogan, 1981) imply that the marginal labor supply function is not the same as the function describing work choice (see Heckman, 1993, for a discussion of this issue in the context of a review of the labor supply literature). Several articles in the labor supply literature have tested the Tobit model for hours of work and have rejected it, not only for men, whose hours are especially clustered, but also for women (Moffitt, 1984; Mroz, 1987).

In the econometric literature, this issue has been partly reflected in discussions of what is known as the Cragg model (Cragg, 1971), which separates the model for y for those with positive y from the model for whether y is positive. The Cragg model is properly considered to be a multiple-equation selection bias model rather than a censored regression model, although the distinction is not important for anything other than nomenclature. In the labor supply literature, the estimation of conditional hours worked functions reflects this same type of model.

Assuming that the object of interest is still the Tobit model, and not a selection bias model – that is, that the model of interest is a single-equation model – we may ask, once again, what the alternatives are to Tobit and how some of these issues may be addressed. As in the models discussed thus far, provided the problem is with the distributional assumption on the error and not with heteroskedasticity or other failure of independence of X and the error term, one solution is to give up on the estimation of β and seek only estimates of $\partial \mathrm{E}(y \mid X)/\partial X$. In the censored regression model in general, without the normality assumption on ε (but still maintaining independence of ε),

$$\mathrm{E}(y \mid X) = \mathrm{Prob}(y > 0 \mid X)\mathrm{E}(y \mid X, y > 0) = [1 - F(-X\beta)]X\beta + \int_{-X\beta}^{\infty} \varepsilon f(\varepsilon \mid X)d\varepsilon$$

$$= g(X), \tag{17}$$

where F and f are the unknown c.d.f. and p.d.f of ε, respectively. A least-squares projec-

tion of y onto X yields a linear approximation to the non-linear curve represented in (17). As before, a NP regression is likely to do a better job in picking up the non-linearities in the curve than least squares.

If β is the object of interest, which it often will be – perhaps more so than in the binary choice case, for here it seems more likely that the continuous sample of the data should allow identification of the index function – a wide variety of econometric methods have been proposed but none has been applied in more than a handful of articles to date, and rarely in labor economics (a recent exception is Chay and Honore, 1998). These include the least absolute deviations estimator (Powell, 1984), the quantile restriction estimator (Powell, 1986a), and the symmetrically trimmed estimator (Powell, 1986b), none of which requires full independence of the error term and hence can accommodate heteroskedasticity. Maintaining independence has led to proposals for a wide variety of additional estimators (Honoré and Powell, 1994 and others). As with the other semi-parametric estimators that have been discussed, there has been insufficient practical experience with these estimators in labor economics for standardized practice to have built up or for very much information to have been gathered on their impact on coefficient estimates in typical applications. Nor have the estimators been incorporated into the major software packages. Given the potential importance of the breakdown of assumptions in the Tobit model, more work in this direction would seem particularly warranted.

As with most semi-parametric estimators in general, most often a parametric assumption on the form of the index function is a maintained assumption. The usual assumption is the standard linear model form $X\beta$, and hence estimates of the features of unknown distributions are implicitly and partly based on deviations from linearity. Little work has been done as well in investigating the interactions of relaxing the linearity of this function with the relaxation of distributional and other assumptions just referred to, either in terms of feasibility, properties, or typical practical performance. As with the binary choice model, there have to be limits to the extent to which linearity can be reduced if independence of the errors is not maintained, because identification of the model can fail completely under a non-parametric specification for the index function combined with arbitrary forms of heteroskedasticity. Unfortunately, the presence of a subset of continuous observations with $y > 0$ does not alter this fundamental problem that also arises in the binary choice model.

2.4. *Sample selection bias model*

The traditional selection bias model in econometrics began with the work of Heckman (1974) on wages and labor supply and was developed, expanded, and elaborated further in a series of papers in the late 1970s by Heckman (1978, 1979), Lee (1979), and others. The literature has two distinct branches, one of which concerns estimation of equations which are observed for only a subsample, either by definition – as in the case of wage rates, which are by definition observed only for those working – or by fortune of data available. This model could be termed the "partial-population" sample selection model

but will here be termed the "sample selection" model for simplicity. The other branch presumes that the total population is available in the data but that there are one or more regressors of interest which take on their values as a result of some type of selection process. The canonical case assumes interest to center on a single dummy variable for some type of treatment and hence these models are often termed "treatment-effect" models. Barnow et al. (1980) drew an analogy between the sample selection model and the treatment-effect model, and proposed estimation techniques for the latter that were based on those developed for the former. But since that time it has become understood that the treatment-effects model admits of a much larger class of estimators, many of which are not applicable to the sample selection model – IV is perhaps the leading case. While there is still some relationship between the two types of models, the literatures have sufficiently diverged that the discussion here will, for space reasons, be restricted entirely to the sample selection model.[19]

The canonical sample selection model can be written as

$$y = X\beta + \varepsilon, \qquad y \text{ observed if } I = 1, \tag{19}$$

$$I^* = Z\delta + v, \tag{20}$$

$$I = 1 \text{ if } I^* \geq 0, \qquad I = 0 \text{ if } I^* < 0, \tag{21}$$

and with the assumption that ε and v are distributed bivariate normal with means zero, variances σ^2 and 1, respectively, and with correlation ρ. The variables I and Z are assumed to be available for the total population. In the sample with observed y, the conditional mean of y is equal to

$$\mathrm{E}(y \mid X, I = 1) = X\beta + \mathrm{E}(\varepsilon \mid X, I = 1) = X\beta + \theta\lambda(Z\delta), \tag{22}$$

where $\theta = \sigma\rho$ and $\lambda(Z\delta) = f(Z\delta)/F(Z\delta)$ is the inverse Mills ratio, and where f and F are the unit normal p.d.f and c.d.f., respectively. Given the result in (22), consistent estimates of β can be obtained either by estimating the two equations in (19)–(21) by maximum likelihood, by a two- step procedure in which probit estimates of (20)–(21) are used to estimate (22) is estimated by least squares (or WLS) using estimates of $Z\delta$ from the first stage, or by a variety of other methods.

Empirical practice in labor economics has seen a decline in the use of these methods, as noted in Section I. This decline in use has a variety of rationales. One is that $\lambda(Z\delta)$ is often

[19] For later developments of the treatment-effects model see Heckman and Robb (1985), Imbens and Angrist (1994), and Manski (1994). The literature is too large to cite many of the developments. The importance of the distinction between the two types of models depends heavily on whether the treatment effect is homogeneous, which is itself related to whether different groups have different equations with separate unobservables. If completely separate equations are specified for the two treatment groups, the model comes closer to the sample selection model. It has also been shown that if the conventional treatment effect coefficient is assumed to be random and a function of the same X variables that appear in the outcome equation, the treatment-effect model separates into the Lee (1979) switching regression model where there are two subpopulations with completely different parameters, which is the same as two separate sample selection models (Björklund and Moffitt, 1987).

highly collinear with X and hence estimates of β tend to be unstable, non-robust, and sensitive to minor changes in the specification of the X and Z vectors. Monte Carlo results of Nelson (1984) show that the standard errors of the elements of β can indeed be very large if the degree of collinearity is high. Other Monte Carlo results show that the inverse Mills ratio is close to linearity over middle ranges of selection probabilities, and exhibits non-linearities that would reduce collinearity with $X\beta$ only in the tails (Leung and Yu, 1996). The argument is usually made for estimation by maximum likelihood as well even though it is more efficient than the two-step method under the model assumptions. A second rationale often mentioned is that the distributional assumption of bivariate normality is unwarranted and may be false and, especially if X and Z coincide, identification of the model is made on the basis of an arbitrary distributional assumption. A third argument often given is that adjustment for sample selection bias does not matter in any case. This rationale is partly in conflict with the first two, for if either collinearity is high or the normality distribution is false, the estimates from the model are not capable of leading to a conclusion one way or the other on the importance of selection bias.[20]

The first two issues are related and have been addressed by the developing semi-parametric literature on sample selection models (Powell, 1994; Vella, 1998). This literature has shown that the bivariate normality assumption can be greatly weakened. A simple relaxation that is partially apparent from (22) already is that all that is really needed for the two-step method is that ν be normally distributed and that ε be linearly related to ν; normality of ε and bivariate normality between the two is not needed. More important, it is clear from (22) that even the normality of ν can be relaxed as long as the joint distribution of ε and ν is independent of X and Z, for in that case the conditional mean of ε depends only on the index function $Z\delta$ which, in turn, depends only on $\text{Prob}(I = 1 \mid Z) = F(Z\delta)$. That is

$$\begin{aligned} \mathrm{E}(y \mid X, I = 1) &= X\beta + \mathrm{E}(\varepsilon \mid X, I = 1) = X\beta + \mathrm{E}(\varepsilon \mid \nu > -Z\delta) = X\beta + h(Z\delta) \\ &= X\beta + h'(p), \end{aligned} \tag{23}$$

where h and h' are unknown functions and $p = \text{Prob}(I = 1 \mid Z\delta)$. Eq. (23) makes no significant distributional assumption on ε and ν (aside from the usual independence assumption from X and Z) and hence can be used as a basis for estimation under relaxed assumptions. Under the same approach as other two-step methods, (23) shows that a first-stage estimate of $Z\delta$ by itself, or of the probability that $I = 1$, if obtainable, can be entered into the equation and used to control for selection bias provided the unknown functions h and h' can be estimated as well.

Approaches along these lines have been elaborated by Gallant and Nychka (1987), Robinson (1988), Choi (1990), Ahn and Powell (1993), and Newey (1988), among

[20] Some of the literature on the robustness of sample selection models is in the treatment-effects literature instead, particularly in the study of the effects of unions and the effects of training programs. Both the nature of the problem and the solution are quite different than in the sample selection model, although there is a similarity in one method of identification (exclusion restrictions) referred to below.

many others. These articles propose that first-stage equations for the probability that $I = 1$ be obtained from semi-parametric or non-parametric methods, thereby reducing or eliminating the parametric assumptions on (20)–(21); that either the estimates of $Z\delta$ or p from the first stage be entered into the second stage and some type of semi-parametric method (kernels, pairwise differences, series estimation, etc.) be used to account for the unknown function h or h^I in the estimation. A somewhat older approach that represents a halfway house between these semi-parametric methods and the conventional parametric, normal model, is one which frees up the bivariate distribution to allow it to be of a form that can capture more types of distribution shapes that the bivariate normal but still maintain a parametric form (e.g., Mroz and Guilkey, 1995). These models are relatively easy to estimate.[21]

In the absence of distributional assumptions, identification of the model requires an exclusion restriction, for it should be clear from (23) that if X and Z coincide, β could not be separated from h or h^I.[22] The source of the collinearity problems that are often experienced in the application of the parametric, normal-based sample selection model are largely the result of either no exclusion restrictions or exclusion restrictions that are weak.

To date the new methods have been very little used and hence their potential in addressing the difficulties associated with the sample selection model have yet to be assessed. One exception is Newey et al. (1990) who applied one version of the semi-parametric method to the classic wage-labor-supply model of Heckman (1974). Interestingly, they found that selection bias adjustment made little difference to estimation of coefficients of the wage equation and that normality could not be rejected. This article may be the source of the view, noted earlier, that selection bias adjustments make little difference. However, much more empirical experience is needed to determine whether this result applies to other groups and datasets, and whether it applies to the enormous range of areas other than the wages of workers where sample selection issues arise before any general conclusion can be reached.[23]

The few applications that have been thus far reported continue to fail to emphasize or explore in depth the issue of exclusion restrictions for identification which come to the fore when distributional assumptions are relaxed. This has also been a problem in past applications of the sample selection model, where exclusion restrictions have been given little attention and have been treated quite casually. The econometric literature has not dealt in great detail with this issue because it is not intrinsically an econometric problem but rather

[21] The Mroz–Guilkey approach is closely related, in turn, to an approach of Heckman and Singer (1984) which was originally applied to hazard models but which is applicable to sample selection models as well. In all these approaches the bivariate distribution of two error terms is assumed to be composed of one error which is discrete multinomial and another is continuous. An issue of theoretical, but as yet unclear practical, importance in this literature is whether these distributions are viewed as the true distributions or only as approximations to the true distributions. The asymptotic distribution of the parameter estimates differs depending on which view is taken.

[22] The intercept cannot be identified in any case under most of these methods but can be estimated by extrapolation, assuming it is of interest.

[23] Vella (1998) presents an example where, unlike Newey et al. (1990), he argues that sample selection adjustments to a wage equation do make a substantive difference.

an economic and empirical problem of finding variables that plausibly affect selection but do not affect y directly. In this respect identification of the sample selection model turns out to have a close affinity to identification in the treatment-effects model, despite the differences in structure of the models noted earlier. In the typical consideration of IV in treatment-effects estimation, the search for instruments which are both (*i*) relevant in the sense of having a strong asymptotic correlation with the endogenous variables holding constant all the other exogenous variables and (ii) which are exogenous have exact parallels in the sample selection model in the search for exclusion restrictions (Z) which are strongly related to the endogenous variable I holding constant X and which are exogenous (independent of ε).[24] In short, then, the solution to identification in the sample selection model, at least if approached through the use of exclusion restrictions, is no more or less difficult than the conventional identification problem through exclusion restrictions that has preoccupied economists since 2SLS was developed and which continues to be a key source of attention in empirical work aiming at the estimation of causal effects. Both in that general literature and in the sample selection literature, an important lesson from much of the empirical work in the last decade is that exclusion restrictions and, more generally, identification cannot be treated cavalierly or in a mechanical fashion; any method which is applied by rote is likely to lead to unsatisfactory results. This is the lesson of the new literature on non-parametric and semi-parametric estimation as well, and it implies that the role of exclusion restrictions should occupy a much more central role in the estimation of sample selection models just as it has come to occupy that role in the treatment-effects literature.

Even given these generalizations, however, there has been much less work in exploring alternative exclusion restrictions in sample selection models compared to treatment-effects models. A body of empirical experience has yet to be built up on the sensitivity of results to different restrictions using the new, less-restrictive methods that have been developed. This should be a topic for research in the future.

3. Conclusions

This survey of econometric methods in labor economics and of recent developments in a few of those methods shows that both practitioners and econometricians are moving in the same direction but without as much contact and interchange as would be fruitful. Empirical work in labor is moving toward less restrictive, more robust, and simpler methods which attempt to isolate and highlight key sources of identification clearly where they can be made the subject of investigation and attention. New developments in econometrics are moving in exactly the same direction but the tools developed there have not spilled over into econometric practice. To do so it is necessary that a body of empirical experience be

[24] Vella (1998) points out that the selection term in (22) can be thought of as a generalized residual from the first-stage regression, which is also closely analogous to IV estimation, for IV can also be formulated by including a first-stage residual in the second-stage equation.

built up so that rules of thumb can be developed, the more useful techniques weeded out from the plethora of those that have been proposed, and incorporated into the commonly used software packages. More work on assessing when and where the relaxation of restrictions makes a difference should be part of this endeavor.

References

Ahn, H. and J. Powell (1993), "Semiparametric estimation of censored selection models with a nonparametric selection mechanism", Journal of Econometrics 58: 3–29.

Aldrich, J. and F. Nelson (1984), Linear probability, logit, and probit models (Sage, Newbury Park).

Amemiya, T. (1981), "Qualitative response models: a survey", Journal of Economic Literature 19: 1483–1536.

Arabmazar, A. and P. Schmidt (1982), "An investigation of the robustness of the tobit estimator to nonnormality", Econometrica 50: 1055–1063.

Ashenfelter, O. and A. Krueger (1994), "Estimates of the economic return to schooling from a new sample of twins", American Economic Review 84: 1157–1173.

Ashenfelter, O. and R. Layard, eds. (1986), Handbook of labor economics, Vols. I and II (North-Holland, Amsterdam).

Barnow, B., G. Cain and A. Goldberger (1980), "Issues in the analysis of selectivity bias", in: E. Stromsdorfer and G. Farkas, eds., Evaluation studies review annual, Vol. 5 (Sage, Newbury Park).

Behrman, J., M. Rosenzweig and P. Taubman (1994), "Endowments and the allocation of schooling in the family and in the marriage market: the twins experiment", Journal of Political Economy 102: 1131–1174.

Berkovec, J. and S. Stern (1991), "Job exit behavior of older men", Econometrica 59: 189–210.

Berkson, J. (1944), "Application of the logistic function to bio-assay", Journal of the American Statistical Association 39: 357–365.

Berkson, J. (1951), "Why I prefer logits to probits", Biometrics 7: 327–339.

Björklund, A. and R. Moffitt (1987), "The estimation of wage and welfare gains in self-selection models," Review of Economics and Statistics 69: 42–49.

Blundell, R. and A. Duncan (1998), "Kernel regression in empirical microeconomics", Journal of Human Resources 33: 62–87.

Bound, J., D. Jaeger and R. Baker (1995), "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak", Journal of the American Statistical Association 90: 443–450.

Brown, C. and R. Moffitt (1983), "The effect of ignoring heteroskedasticity on estimates of the tobit model", Technical working paper no. 27 (NBER, Cambridge, MA).

Choi, K. (1990), "The semiparametric estimation of the sample selection model using series expansion and the propensity score", Unpublished manuscript (University of Chicago).

Cogan, J. (1981), "Fixed costs and labor supply", Econometrica 49: 945–964.

Cragg, J. (1971), "Some statistical models for limited dependent variables with application to the demand for durable goods", Econometrica 39: 829–844.

Domenich, T. and D. McFadden (1995), Urban travel demand (North-Holland, Amsterdam).

Finney, D.J. (1947), Probit analysis (Cambridge University Press, Cambridge).

Fisher, R.A. (1925), Statistical methods for research workers (Oliver and Boyd, Edinburgh).

Gallant, R. and D. Nychka (1987), "Semi-nonparametric maximum likelihood estimation", Econometrica 55: 363–390.

Geweke, J. and M. Keane (1999), "Mixture of normals probit models", in: C. Hsiao, K. Lahiri, L. Lee and H. Pesaran, eds., Analysis of panels and limited dependent variable models: an edited volume in honor of G.S. Maddala (Cambridge University Press, Cambridge, UK) in press.

Geweke, J., M. Keane and D. Runkle (1994), "Alternative computational approaches to inference in the multinomial probit model", Review of Economics and Statistics 76: 609–632.

Goldberger, A. (1964), Econometric theory (Wiley, New York).

Goldberger, A. (1983), "Abnormal selection bias", in: S. Karlin, T. Amemiya and L. Goodman, eds., Studies in econometrics, time series, and multivariate statistics (Academic Press, New York).

Hajivassiliou, V. (1993), "Simulation estimation methods for limited dependent variable models", in: G.S. Maddala, C.R. Rao and H.D. Vinod, eds., Handbook of statistics, Vol. 11 (North-Holland, Amsterdam).

Hajivassiliou, V. and P. Ruud (1994), "Classical estimation methods for limited dependent variable models using simulation", in: R. Engle and D. McFadden, eds., Handbook of econometrics, Vol. IV (North-Holland, Amsterdam).

Härdle, W. and O. Linton (1994), "Applied nonparametric methods", in: R. Engle and D. McFadden, eds., Handbook of econometrics, Vol. IV (North-Holland, Amsterdam).

Hausman, J. (1980), "Effects of wages, taxes, and fixed costs on women's labor force participation", Journal of Public Economics 14: 161–194.

Hausman, J. and D. McFadden (1984), "Specification tests for the multinomial logit model", Econometrica 52: 1219–1240.

Heckman, J. (1974), "Shadow prices, market wages, and labor supply", Econometrica 42: 679–693.

Heckman, J. (1978), "Dummy endogenous variables in a simultaneous equation system", Econometrica 47: 931–959.

Heckman, J. (1979), "Sample selection bias as a specification error", Econometrica 47: 153–161.

Heckman, J. (1993), "What has been learned about labor supply in the past twenty years?" American Economic Review 83: 116–121.

Heckman, J. (1997), "Instrumental variables", Journal of Human Resources 32: 441–462.

Heckman, J. and R. Robb (1985), "Alternative methods for evaluating the impact of interventions", in: J. Heckman and B. Singer, eds., Longitudinal analysis of labor market data (Cambridge University Press, Cambridge, UK).

Heckman, J. and B. Singer (1984), "A method for minimizing the impact of distributional assumptions in econometric models for duration data", Econometrica 52: 271–320.

Heckman, J. and J. Snyder (1996), "Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators", Working paper 5785 (NBER, Cambridge, MA).

Honoré, B. and J. Powell (1994), "Pairwise difference estimators of censored and truncated regression models", Journal of Econometrics 64: 241–278.

Horowitz, J. (1993a), "Semiparametric estimation of a work-trip choice model", Journal of Econometrics 58: 49–70.

Horowitz, J. (1993b), "Semiparametric and nonparametric estimation of quantal response models", in: G.S. Maddala, C.R. Rao and H.D. Vinod, eds., Handbook of statistics, Vol. 11 (North-Holland, Amsterdam).

Hurd, M. (1979), "Estimation in truncated samples when there is heteroskedasticity", Journal of Econometrics 11: 247–258.

Imbens, G. and J. Angrist (1994), "Identification and estimation of local average treatment effects", Econometrica 62: 467–475.

Keane, M. (1992), "A note on identification in the multinomial probit model", Journal of Business and Economic Statistics 10: 193–200.

Keane, M. (1993), "Simulation estimation for panel data models with limited dependent variables", in: G.S Maddala, C.R. Rao and H.D. Vinod, eds., Handbook of statistics, Vol. 11 (North-Holland, Amsterdam).

Keane, M. (1994), "A computationally practical simulation estimator for panel data", Econometrica 62: 95–116.

Keane, M. and R. Moffitt (1998), "A structural model of multiple welfare program participation and labor supply", International Economic Review 39: 553–589.

Lee, L. (1979), "Identification and estimation in binary choice models with limited (censored) dependent variables", Econometrica 47: 977–996.

Lerman, S. and C. Manski (1981), "On the use of simulated frequencies to approximate choice probabilities", in:

C. Manski and D. McFadden, eds., Structural analysis of discrete data with econometric applications (MIT Press, Cambridge, MA).

Leung, S. and S. Yu (1996), "On the choice between sample selection and two-part models", Journal of Econometrics 72: 107–128.

McFadden, D. (1974), "Conditional logit analysis of qualitative choice behavior", in: P. Zarembka, ed., Frontiers of econometric analysis (Academic Press, New York).

McFadden, D. (1981), "Econometric models of probabilistic choice", in: C. Manski and D. McFadden, eds., Structural analysis of discrete data with econometric applications (MIT Press, Cambridge, MA).

McFadden, D. (1989), "A method of simulated moments for estimation of discrete response models without numerical integration", Econometrica 57: 996–1026.

Maddala, G.S. (1995), "Specification tests in limited dependent variable models", in: G.S. Maddala, P.C.B. Phillips and T.N. Srinivasan, eds., Advances in econometrics and quantitative economics (Blackwell, Oxford, UK).

Manser, M. (1999), "Existing labor market data: current and potential research uses", in: J. Haltiwanger, M. Manser and R. Topel, eds., Labor statistics measurement issues (The University of Chicago Press, Chicago, IL) in press.

Manski, C. (1975), "Maximum score estimation of the stochastic utility model of choice", Journal of Econometrics 3: 205–228.

Manski, C. (1988), "Identification of binary response models", Journal of the American Statistical Association 83: 729–738.

Manski, C. (1991), "Regression", Journal of Economic Literature 29: 34–50.

Manski, C. (1994), "The selection problem", in: C. Sims, ed., Advances in econometrics (Cambridge University Press, Cambridge, UK).

Manski, C. and S. Thompson (1986), "Operational characteristics of maximum score estimation", Journal of Econometrics 32: 85–108.

Matzkin, R. (1992), "Nonparametric and distribution-free estimation of the binary crossing and binary choice models", Econometrica 60: 239–270.

Melenberg, B and A. Van Soest (1996), "Parametric and semi-parametric modeling of vacation expenditures", Journal of Applied Econometrics 11: 59–76.

Moffitt, R. (1984), "The estimation of a joint wage-hours labor supply model", Journal of Labor Economics 2: 550–566.

Mroz, T. (1987), "The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions", Econometrica 55: 765–800.

Mroz, T. and D. Guilkey (1995), "Discrete factor approximations for use in simultaneous equation models with both continuous and discrete endogenous variables", Working paper no. 95-02 (Department of Economics, University of North Carolina, Chapel Hill, NC).

Nelson, F. (1984), "Efficiency of the two-step estimator for models with endogenous sample selection", Journal of Econometrics 24: 181–196.

Newey, W.K. (1988), "Two-step series estimation of sample selection models", Unpublished manuscript (Princeton University).

Newey, W., J. Powell and J. Walker (1990), "Semiparametric estimation of selection models: new results", American Economic Review 80: 324–328.

Pagan, A. and F. Vella (1989), "Diagnostic tests for models based on individual data: a survey", Journal of Applied Econometrics 4: S29–S59.

Pakes, A. and D. Pollard (1989), "Simulation and the asymptotics of optimization estimators", Econometrica 57: 1027–1058.

Pencavel, J. (1986), "Labor supply of men", in: O. Ashenfelter and R. Layard, eds., Handbook of labor economics (North-Holland, Amsterdam).

Powell, J. (1984) "Least absolute deviations estimation for the censored regression model", Journal of Econometrics 25: 303–325.

Powell, J. (1986a) "Censored regression quantiles", Journal of Econometrics 32: 143–155.
Powell, J. (1986b) "Symmetrically trimmed least squares estimation of tobit models", Econometrica 54: 1435–1460.
Powell, J. (1994), "Estimation of semiparametric models", in: R. Engle and D. McFadden, eds., Handbook of econometrics, Vol. IV (North-Holland, Amsterdam).
Quandt, R. (1956), "A probabilistic model of consumer behavior", Quarterly Journal of Economic Behavior 70: 507–536.
Robinson, P.M. (1993), "Root-N consistent semiparametric regression", Econometrica 56: 931–954.
Staiger, D. and J. Stock (1997), "Instrumental variables regression with weak instruments", Econometrica 65: 557–586.
Stafford, F. (1986), "Forestalling the demise of empirical economics: the role of microdata in labor economics research", in: O. Ashenfelter and R. Layard, eds., Handbook of labor economics (North-Holland, Amsterdam).
Stern, S. (1997), "Simulation-based estimation", Journal of Economic Literature 35: 2006–2039.
Theil, H. (1981), Principles of econometrics (Wiley, New York).
Thurstone, L. (1927), "A law of comparative judgment", Psychological Review 34: 273–286.
Tobin, J. (1958), "Estimation of relationships for limited dependent variables", Econometrica 26: 24–36.
Vella, F. (1998), "Estimating models with sample selection bias", Journal of Human Resources 33: 127–169.

Chapter 25

INSTITUTIONS AND LAWS IN THE LABOR MARKET

FRANCINE D. BLAU

Cornell University and NBER

LAWRENCE M. KAHN

Cornell University

Contents

Abstract 1400
JEL codes 1400
1 Introduction 1400
2 Scope of the chapter 1403
3 Why are there labor market institutions? 1404
4 Theoretical overview: recent developments in analyzing labor market institutions, wage levels, wage dispersion, and employment 1407
4.1 Union-management bargaining: beyond the monopoly union-efficient bargaining dichotomy 1408
4.2 The impact of centralization of wage-setting 1409
4.3 Employment protection, employment and wages 1411
4.4 The relationship between wage-setting, centralization and social policies 1413
5 Wage centralization and macroeconomic performance 1414
6 Wage-setting institutions and wage inequality 1416
6.1 Collective bargaining institutions and overall wage inequality 1417
6.2 Wage-setting institutions and the relative pay of particular groups: women 1429
6.3 Wage-setting institutions and the relative pay of particular groups: minimum wage laws 1433
7 Responses to labor market institutions 1434
7.1 Employment and wage-setting institutions 1435
7.2 Labor market flexibility and employment responses 1443
7.3 Countervailing institutional responses to adverse labor market consequences of institutions 1449
8 Conclusions 1453
References 1455

Handbook of Labor Economics, Volume 3, Edited by O. Ashenfelter and D. Card

Abstract

This chapter examines the impact of wage-setting institutions and government policies on wages and employment, focusing on the OECD countries. There is considerable evidence that centralized collective bargaining, minimum wages and antidiscrimination policies raise the relative wages of the low paid. Evidence of the impact of these institutions and other policies such as mandated severance pay, advance notice or unemployment insurance is more mixed with some studies finding negative employment effects while others do not. This may reflect the adoption by many OECD countries of off-setting policies, such as public employment, temporary employment contracts and active labor market programs, which, while they may have reduced the adverse relative employment effects of their less flexible labor market institutions on the low skilled, appear not to have prevented high overall unemployment.

JEL codes: J16; J31; J38; J51; J65

1. Introduction

This chapter is concerned with the impact of institutions on the labor market. What we mean by institutions, in this context, is the system of laws, programs, and conventions that can impinge on labor market behavior and cause the labor market to function differently from a spot market. Over the last 10 years there has been an explosion of research on the economic impact of such institutions. This activity has been fueled by a variety of factors, including economic developments in the 1980s and 1990s, especially international differences in labor market performance; methodological innovations in empirical labor economics; and the increasing availability of large-scale microdata sets for many countries. Each of these factors in turn leads us to take a comparative focus in this chapter on institutions and laws in the labor market.

Diverging labor market performance across countries in the 1980s and 1990s has caused researchers and policy makers to examine labor market institutions in an attempt to learn "what works." As may be seen in Table 1, the US had higher unemployment than most other OECD countries in the early 1970s, but by the 1980s the situation had reversed itself and the US had become a relatively low unemployment country. Although the US had a smaller population than Europe, it generated 20 million net additional jobs between 1975 and 1985 compared to only 2 million for the European OECD countries (Freeman, 1994). European unemployment stayed stubbornly high into the mid-1990s, averaging over 9%, while the US rate fell toward 5%. The stark contrast between the US record of strong job creation and relatively low unemployment and the European experience of sluggish job growth and persistently high unemployment has led many Europeans to reexamine their labor market institutions of administered wages and legislated job protections and social benefits, in comparison to the considerably more flexible, less regulated labor market of the United States (OECD, 1994b).

Table 1
Standardized unemployment rates in OECD countries, 1973, 1984 and 1995[a]

	1973	1984	1995
Austria	1.1	3.8	4.3
Belgium	2.7	14.0	9.4
Denmark	1.7	8.5	10.0
Finland	2.3	6.1	17.1
France	2.6	9.7	11.6
Germany (W)	0.7	8.5	6.7
Ireland	5.7	15.5	12.9
Italy	6.2	10.2	12.2
Netherlands	2.2	14.0	6.5
Norway	1.5	3.0	4.9
Spain	2.5	20.1	22.7
Sweden	2.8	3.1	9.2
Switzerland	0.4	1.1	3.3
UK	3.0	13.0	8.7
Australia	1.3	8.9	8.5
Canada	5.5	11.2	9.5
Japan	1.3	2.7	3.1
New Zealand	0.2	5.7	6.3
Non-US average (unweighted)	2.4	8.8	9.3
US	4.8	7.4	5.5

[a] Source: Compiled from Freeman (1988, p. 70), Nickell (1996, p. 13), OECD (1983, p. 23) and OECD (1996, p. 198).

On the other hand, while the US has achieved much greater employment growth than other OECD countries, American real wages have been rising more slowly (Freeman, 1994). Moreover, the rising wage inequality that affected many advanced nations during the 1980s and 1990s was especially pronounced in the US and, in several OECD countries, inequality hardly increased at all. The US also appears to be the only country with sharply *falling* real wages of less-skilled workers during this period (Freeman and Katz, 1995). While all of these countries have likely been affected by similar changes in technology and by the growth of international trade, they have very different labor market institutions. On a priori grounds, then, labor market institutions appear to be a promising explanation for at least some of these differences in labor market performance.

While substantive policy issues have undoubtedly been responsible for the lion's share of increased interest in the impact of alternative labor market institutions, methodological issues have also played an important role. In particular, an exciting recent development in labor economics is the "natural experiment" approach to hypothesis testing. By this we mean a search for exogenous variation in key explanatory variables in an attempt to

identify their causal effects. Frequently a policy change or difference provides the source of this variation.[1]

The natural experiment approach seems tailor-made for a research design which uses international comparisons to identify the impacts of labor market institutions. In particular, the OECD countries are similar in many respects, including their educational systems, forms of government, living standards and levels of economic development, at least in comparison to the rest of the world. However, their labor market institutions differ greatly and have changed over time at different rates. Such variations allow for sharp tests of the impact of institutions in a "differences-in-differences" framework. For example, while Canada and the US are very similar in most important respects, unionization has fallen sharply in the US but remained fairly stable in Canada (Card and Freeman, 1993). Thus the US and Canada provide a nice comparison for evaluating the impact of this institutional change. Further, cross-country comparisons may enable us to obtain considerably wider variation in some basic explanatory variables than occurs within a country over time or across areas. For example, in several European countries, collective bargaining agreements cover at least 90% of workers, while in the US, the figure is below 20% (OECD, 1994a). Such large differences enhance the probability of empirically detecting the effects of the relevant explanatory variables. Moreover, the recent development of large scale microdata bases for a large number of countries permits researchers to control for many other influences on labor market outcomes, such as human capital characteristics, in an attempt to focus on the impact of institutional differences.

A large body of research contrasts the US experience to that of other advanced countries because the US stands at one extreme in terms of its distinctively flexible, less regulated labor market and because the international differences in outcomes are largest between the US and other countries. Thus, an organizing theme of this chapter is to examine the role of institutions by comparing various US outcomes to those in other OECD countries. Yet there are other instructive comparisons as well. For example, among the culturally similar Scandinavian countries, institutions have not always changed at the same rate or even in the same direction, and these comparisons, like the case of the US and Canada, provide a good natural experiment for examining the effects of institutions.

The plan of this chapter is as follows. In the next section, we define the scope of issues examined and countries studied. We then consider theoretical issues in estimating the impact of institutions. These include explanations for the emergence of particular institutions and theoretical expectations for their labor market impacts. We then review empirical research on the effects of institutions on wage levels, wage distributions, and employment. We conclude with some observations about the coexistence of different labor market regulatory regimes in a world economy.

[1] Some US examples include the impact of compulsory school attendance laws (Angrist and Krueger, 1991); changes in minimum wage laws (Card and Krueger, 1995); and changes in Federal policy toward maternity care coverage in company sponsored health insurance plans (Gruber, 1994).

2. Scope of the chapter

In the interests of manageability, we have placed certain limits on the types of institutions examined and the contexts in which they are studied. We focus primarily on the impact of direct attempts to change market outcomes with respect to pay-setting and firms' utilization of labor. These institutions and laws directly regulate what workers and firms are permitted to do in setting wages and allocating labor. In addition, there are a number of policies designed to ameliorate the outcomes produced by unrestricted markets that indirectly affect worker-firm interactions. Such policies may also have important impacts on wages and employment. Thus we consider several of them as well.

Examples of direct intervention in the wage-setting process include collective bargaining agreements, as well as policy interventions regulating wage determination. Government policies include rulings extending the terms of collective bargaining contracts to workers not initially covered by the agreement, anti-discrimination policy as it relates to pay, minimum wage laws, and the behavior of the government as an employer. In addition to regulating wages, governments in each of the countries we examine have placed some limits on the unrestricted right of firms to fire workers. For example, most European countries require severance pay and advanced notice in the event of layoffs. Policies having a more indirect effect on labor utilization and wage determination include payroll taxes, unemployment insurance, industrial subsidies, and active labor market policies, including public employment and training programs, and we review important evidence on these as well. However, we do not examine in detail every intervention that could conceivably affect wages or employment. So, for example, we do not discuss in detail the impact of international differences in income tax systems, policies toward imports, occupational safety and health, or regulations governing firm entry or rate of return. This decision reflects considerations of space as well as a paucity of internationally comparable data on the impact of such policies.

As our earlier discussion suggests, comparing outcomes across culturally similar countries provides a very attractive research design for evaluating the impact of institutions. For this reason, we limit the countries considered here to the roughly 20 nations in the OECD. Not only does this limitation keep our review manageable, it also permits us to utilize the similarity in educational levels, technology, living standards and cultures among these countries as de facto controls in examining the effects of institutions. It should be noted that, in our discussion below, we frequently use the term "European" in referring to institutional arrangements or outcomes that are highly characteristic of that region, although some OECD countries outside of Europe, notably Australia and New Zealand, share the major features of this model.

While our focus on the OECD nations provides a considerable degree of comparability across countries along a number of important dimensions, significant non-comparabilities remain, some of which reflect labor market institutions themselves. For example, it is likely that many people who would be classified as out-of-the labor force in the US would be classified as long-term unemployed in Europe because they would receive long-term

unemployment benefits there which are not available in the US. This may be expected to drive up the incidence of long-term unemployment in Europe compared to the US, all else equal. While this is a serious concern, we do not believe it unduly affects our examination of the impact of institutions in this chapter. This is because the "differences-in-differences" methodology employed by many, though certainly not all the studies we consider, nets out the impact of such factors by comparing within country changes over time or by making comparisons across groups within a country (or a combination of these designs). In addition, precisely because of the importance of institutions and cultural factors in influencing measured unemployment rates even within countries, we tend to prefer studies which focus on employment-to-population ratios. With some exceptions, the results of studies using the latter concept are generally consistent with those focusing on the unemployment rate. Nonetheless, the way institutions and other factors affect labor market concepts like unemployment or disability as another example would make for an interesting research question which we hope will receive greater attention now that international comparisons are becoming more prevalent.

3. Why are there labor market institutions?

Before analyzing the impact of institutions on the labor market it is worth considering the reasons for their existence in the first place. This issue is of interest in its own right, since an understanding of where the demand for institutions comes from will help us predict likely institutional responses to economic developments. For example, given indexation costs, a large increase in inflation uncertainty is likely to raise the demand for wage indexation on the part of risk-averse workers (Ehrenberg et al., 1983). Similarly, declines in inflation uncertainty are expected to reduce workers' resistance to limiting systems of indexation, as appears to have occurred in Italy the mid-1980s (Erickson and Ichino, 1995).

A second important motivation for considering the reasons for the emergence of institutions is related to economic methodology. An observed empirical association between institutions and outcomes, such as unemployment rates or wage inequality, may indeed reflect a true causal relationship. However, it is also possible that the causation runs from the outcomes to the establishment of the institutional arrangements or that the same set of underlying factors led both to the formation of the institutions and to the observed economic outcomes. In these latter cases, we may overestimate the impact of institutions on economic outcomes. For example, as we shall see, declining unionization has been credited by some with causing a portion of the observed increase in wage inequality in particular countries. However, if such factors as international trade and technological change contributed to the reduction in unionization, then these authors are attributing too much of the increase in inequality to deunionization. Nonetheless, even granting that institutions may be endogenous, we still expect them to play an important role in moderating the impact of economic forces. This is one source of the demand for institu-

tions in the first place. Of course it is also that case that market forces may reassert themselves after an institutional arrangement is implemented or modified. For example, the US minimum wage may be viewed as an institution that has relatively little effect on most labor markets because it is so low. The extent to which market forces override the impact of institutions in particular cases is an empirical question.

Two approaches to the question of the emergence of institutions have been proposed in recent years. First, it has been argued by some that labor market regulations can correct market failures, usually related to imperfect information. Such institutional interventions may increase economic efficiency by changing the outcome that would have resulted from the operation of market forces. Second, others have taken an explicit political economy approach to explaining the emergence of institutions. This entails identifying politically powerful groups and attempting to understand what is in their economic self-interest. We now provide some examples of these two kinds of explanations.

While there may be other types of labor market failures that could potentially be remedied by government intervention, the ones that are most prominent in the literature involve job security, wage-setting, or job search. First, firms can benefit from offering workers job security because such an arrangement is likely to make workers more willing to undertake firm-specific training investments for which the returns would otherwise be uncertain (Hashimoto, 1990). However, if only one firm in a labor market offers job security (implicitly or explicitly), workers who would have otherwise expected to be discharged from their jobs will be among those most attracted to that firm. If there is also asymmetric information, that is if workers have better information about their likely performance than firms do, then the resulting adverse selection problem is likely to reduce the willingness of firms to offer job security. Just as in insurance markets, we may get a suboptimal level of job security. However, laws mandating job security, or raising the cost of discharging workers, can help solve the adverse selection problem and therefore raise economic efficiency (Levine and Tyson, 1990). The same reasoning has been applied to government mandated parental leave policies. Without such mandates, individual firms offering leave will attract the workers most likely to take leave (Ruhm and Teague, 1997).

A second insurance-related motivation for institutions has been used to provide an efficiency rationale for wage-equalizing mechanisms, such as highly centralized collective bargaining structures, that impose a low level of wage inequality on major portions of the labor market. If, before entering the labor market, risk averse individuals are uncertain about how the market is going to reward their human capital, they may have a demand for income insurance. However, private insurance markets will not supply such a product for the usual reasons of adverse selection and moral hazard. Ljungqvist (1995) and Agell and Lommerud (1992) interpret wage leveling as a form of income insurance, implying that institutions leading to such an outcome may raise economic welfare.

An additional potentially welfare-enhancing effect of wage equalization could be realized if there would have been large industry wage differentials in the absence of wage-setting intervention. As Bulow and Summers (1986) argue, when some industries pay efficiency wages, labor may be misallocated: marginal revenue products of labor for

identical workers will vary across sectors of the economy. An encompassing union wage policy of equalization across industries can eliminate this source of inefficiency. Teulings and Hartog (1998) argue that this reasoning characterizes economies where wages are set at a highly centralized level such as in the Scandinavian countries. However, to the extent that there are other sources of interindustry differentials such policies can also create inefficiencies by eliminating wage differentials that would otherwise encourage people to move to sectors where labor is scarce or to invest in on-the-job training. Finally, active labor market policies such as training or relocation allowances may overcome the results of market failures in matching workers and employers. Unemployment insurance can also be viewed this way to the extent that it helps workers find better matches.

While these theories remind us that institutional interventions can potentially increase economic welfare on efficiency, as well as on equity, grounds, they do not yield sharp predictions about where and when particular institutions are most likely to emerge. For example, we might expect government intervention regulating job security provisions to be more likely the worse the adverse selection problem facing private firms in the absence of government action. In general, such problems are more severe the greater the heterogeneity of the labor force with respect to productivity and the more difficult it is for firms to learn about the productivity of workers prior to hiring them. Thus, we might expect greater demand for job security-enhancing policies on efficiency grounds in countries with more heterogeneous workers and with poorer labor market information about their ability. But, there may be less consensus in favor of job-security policies in such settings than where the electorate is more homogeneous. Similarly, controlling for risk aversion, the demand for wage-leveling should be greater the larger is the ex ante uncertainty regarding market wages for an individual. But finding a coalition in support of such policies in economies with widely divergent labor market outcomes for seemingly similar individuals may be difficult.

These examples imply that issues of political economy can soften the predictions of efficiency-based models for the formation of institutions. A political economy approach, with some reasonable assumptions, can help us understand the growth and demise of certain labor market rules and laws. Saint-Paul (1996) uses such a framework to analyze what he claims to be the European resistance to reforming the traditional labor market institutions of generous unemployment insurance (UI), high minimum wages, and high firing costs, even in the face of persistent 10% unemployment levels. His basic framework in studying laws and institutions is to focus on the "decisive voter" in the democratic systems of the European OECD countries. For example, to the extent that the employed are more numerous and well-organized than the non-employed, policies benefiting the former are likely to be adopted even if they result in persistent high unemployment. High mandated firing costs are an example of a law that protects incumbents ("insiders") at the expense of those without jobs ("outsiders"). More generally, insiders may push for very high wage levels, as long as their own jobs are protected. On the other hand, Saint-Paul (1996) predicts that the larger the group of outsiders, the more likely labor market reforms that increasing firm flexibility will be enacted. For example, in the early 1990s, in Spain,

about 20% of the labor force was unemployed, and 33% of the employed were on fixed duration contracts (Bentolila and Dolado, 1994). Since nearly half of the labor force could be considered outsiders, one might suppose that this country was ripe for labor market reforms benefiting these groups. Below, we consider research which attempts to evaluate the impact of policies designed to increase employment flexibility in Spain, including the issue of which groups have benefited from these changes.

Like the economic efficiency framework, the political economy approach also may not yield airtight predictions. This may be due, in part, to the problems in aggregating preferences and the changing identity of the groups (today's outsiders may anticipate being tomorrow's insiders). And many of the features of European insiders also characterize Americans employed at high wages, implying that something more is needed if one is to explain why the US has so much more labor market flexibility than other OECD countries. Yet the political economy analysis forces us to focus on who gains and who loses from specific policies and can in many instances yield interesting insights about why governments make the decisions they do.

Ultimately, the US-European differences in labor market flexibility likely reflect the same factors that have resulted in such a low rate of unionization in the United States. The origins of this difference date back to the 19th century when, it has been argued, individuals perceived much greater opportunity for upward social mobility and wealth accumulation in the US than in Europe (Pelling, 1960). If these perceptions were in fact held, then it is easy to see why collectivist policies of social democratic trade unionism and welfare capitalism took hold to a lesser extent in the United States than in Europe. And this may well be the essential reason why the US labor market remains so much more flexible.

4. Theoretical overview: recent developments in analyzing labor market institutions, wage levels, wage dispersion, and employment

As in other areas of labor economics, the focus of much of the recent research on institutions and laws in the labor market has been empirical. However, there have also been some important theoretical developments regarding at least three aspects of interventions in the labor market. First, economists have refined early models of union-management interaction that were primarily concerned with whether collective bargaining agreements left firms on their labor demand curves. These new models reflect developments in game theory and several interesting new implications have resulted. Second, economists have built models concerning the impact of centralization of wage-setting on union-management bargaining behavior. This work asks, for example, what difference it makes to have encompassing unions and employer federations relative to individual union-firm bargaining units. The impact of different forms of collective bargaining on aggregate wage pressure and employment levels, as well as wage inequality, have been addressed here. Third, the employment and wage effects of employment protection have been analyzed from a theoretical perspective. One question addressed here is the expected effect of high

severance pay on average employment levels. These types of theoretical analyses lead to some interesting insights that have been tested on a variety of data from the OECD countries; and we survey this empirical work in the following section.

4.1. Union-management bargaining: beyond the monopoly union-efficient bargaining dichotomy

Earlier theoretical research on collective bargaining behavior focused on questions of whether such bargains allowed firms to remain on their labor demand curves (the "monopoly union" model) or whether wages and employment were set simultaneously ("the efficient bargaining" model).[2] This literature concluded that under either approach, unions are expected to raise the relative wages of their members, but that employment effects were ambiguous – negative in the monopoly union model but positive or zero in most efficient bargaining models.

The negative effect of unions on employment in the monopoly union model is a direct consequence of the assumption that employers are allowed to be on their labor demand curves. In the "strong" efficient bargaining model, union members are assumed to be risk neutral. Efficient (Pareto-optimal) bargains in this case maximize labor and management's joint surplus by calling for the competitive (i.e., efficient) level of employment and giving the union a share of the proceeds. In this case, unions do not affect employment levels, but do raise wages. However, when union members are risk averse and when there can be no side payments from employed to unemployed union members, then efficient contracts will in general raise wages *and* employment compared to the non-union outcome. The possibility that unions could actually raise employment at the micro level, or at least not reduce it, was a major development and led to a rethinking of the potential impact of trade unions.[3]

In recent years, some important refinements to the basic monopoly union-efficient bargaining dichotomy have been added that build on recent developments in game theory. Specifically, two perhaps unrealistic or unexplored aspects of these earlier models have been examined. First, in the monopoly union model, it was originally assumed that unions could costlessly impose any wage they wanted on management, knowing the quantity of labor that would be demanded as a result. This approach ignores the fact that there is wage bargaining, even when firms unilaterally control the quantity of labor demanded. An important modification, then, has been to assume that there is Nash bargaining over the wage even in monopoly union models. In such models, the status quo utilities of workers and firms (i.e., what each can achieve in the event negotiations continue without an agreement) will directly affect the bargained wage outcome (Moene, 1988; Manning, 1987, 1994).

[2] For a summary, see Farber (1986).

[3] However, even if there is efficient bargaining at the micro level, in a general equilibrium context, overall employment may not rise as a result of collective bargaining (Layard and Nickell, 1990).

This theoretical result implies that the set of factors that influence union wages in the monopoly union model is richer than previously supposed. Specifically, in the original model, the only factors affecting wages were the slope of the labor demand curve and the union's tradeoff between wages and employment. In the Nash bargaining approach, the status quo income of the firms also directly affects wages. The status quo income of union members, e.g., wages on alternative jobs or UI benefits, already affected wages since it is likely to influence the union's tradeoff between wages and employment.

The second refinement of the traditional union-management bargaining models was to question how the efficient bargaining model can be enforced, since it puts firms off their labor demand curves. If firms are not closely monitored, they may attempt to move back onto the demand curve, and it may be difficult for unions to determine whether or not this has occurred. This setup was ripe for analysis as a repeated game, in which the possibility of punishment could enforce the cooperative outcome of efficient bargaining. The monopoly union solution is analogous to the non-cooperative outcome in a prisoner's dilemma game, where the firm moves to its demand curve and the union imposes the monopoly union wage. An immediate prediction from the theory of repeated games is that the lower the discount rate, the more likely the cooperative solution is to be an equilibrium (Espinosa and Rhee, 1989), although the non-cooperative outcome is still always also an equilibrium. This is the case because the lower the discount rate, the less likely that the present value of the short-term gains of defecting from the cooperative solution will exceed the present value of the long-term losses. And some US evidence indicates that unions are more likely to enter into cooperative labor agreements in settings with long time horizons, as predicted by the theory of repeated games (Kahn, 1993).

4.2. The impact of centralization of wage-setting

Formal examination of the impact of wage centralization began in 1988 with a very influential paper by Calmfors and Driffill (1988) which was primarily concerned with macroeconomic performance. The authors attempted to explain why some collective bargaining systems led to wage restraint and high employment levels, while other systems yielded high real wages but low employment levels. This work and the extensions that followed used the monopoly union framework in which employment is determined unilaterally by firms after wage bargaining; hence, the presumed negative relationship between real wage levels and employment followed.

The key insight of the Calmfors and Driffill (1988) approach was to note that the level at which negotiations take place will greatly influence the wage-bargaining stance taken by unions and management. Centralization refers to the degree to which coalitions are created across unions and across firms or industries. At one extreme, consider the most decentralized form of collective bargaining – enterprise bargaining between one union and part or all of one firm. This type of bargaining characterizes the US more than other countries with the possible exception of Canada.[4] If the industry in question is competitive, there

will be almost no scope for a union to raise its members' relative wages, so we will observe wage restraint and high employment levels.

Making negotiations more centralized has two opposing effects on union wage policy. On the one hand, bringing more firms into the bargaining unit lowers the effective elasticity of demand for labor. For example, if instead of organizing only one firm in the industry, the union is able to induce all firms to join the bargaining unit, then union wage policy is likely to become more aggressive, as the employment losses caused by a given wage increase are reduced. On the other hand, the more workers that are included in the bargaining unit, the better able the union is to internalize what would have been externalities had bargaining been less centralized. For example, a union in a single-firm bargaining unit is unlikely to take into account the effects of higher union-negotiated wages on other workers (or even its own members) through higher prices or higher taxes to finance larger unemployment insurance payouts. In contrast, when an encompassing union signs a contract with a large employer federation covering all workers in all industries, the price effects of higher wages directly lower the union members' real wages and the taxes that pay for higher UI benefits will come out of union members' incomes.[5] In this situation, the union will practice wage restraint because the price and tax reactions to high wages hurt the union members themselves. Calmfors and Driffill's (1988) model thus predicts wage restraint for encompassing unions as well as completely decentralized union-management bargaining pairs. The intermediate case is the worst from a macroeconomic point of view: enough centralization to assure the union's ability to raise wages without much job loss, but not enough centralization to induce the union to take into account the price and tax consequences of its wage bargaining.

This basic framework has been extended to include analyses of monopolistic competition, international trade, and insider-outsider issues (for a summary, see Calmfors, 1993). These considerations may modify the inverted U-shaped relationship between wage restraint and centralization. For example, under monopolistic competition, even unions bargaining at the level of the individual firm can raise wages, but the arguments for a higher wage under industry-level bargaining than under firm-level bargaining remain: consumers find it more difficult to substitute across industries than among firms within an industry (Layard et al., 1991). Trade can counterbalance the postulated effects of centralization. With an open economy, foreign competition will restrain even industry-wide monopoly unions (Danthine and Hunt, 1994). In the limit, in a small, open economy with world markets for each good it produces, centralization will make no difference at all (Calmfors, 1993).

Bargaining became increasingly decentralized during the 1980s and 1990s across a number of countries, including Australia, Germany, Italy, the US, the UK and Sweden

[4] Of course, individual, non-union wage setting is even more decentralized than enterprise unionism, and the US, with its low union coverage, is highly decentralized along this dimension as well.

[5] This is clearly true if UI benefits are financed out of income tax revenues. In the more likely case in which payroll taxes pay for UI benefits, union members' incomes will be lowered to the extent that some of the taxes are shifted to wages, a likely outcome (see below).

(Katz, 1993). This development, particularly the end of economy-wide frame agreements in Sweden in 1983, inspired Freeman and Gibbons' (1995) game theoretic model of wage centralization. In their view, there are four potential sets of players in the process that determines the degree to which wages are set at the national level: local unions, individual firms, union federations and employer federations. In deciding how much local wage variation to allow through wage drift, union federations weigh the costs and benefits of keeping wages centrally determined. The benefits involve the ability to restrain inflationary wage hikes and tax increases to pay for higher unemployment benefits; the costs involve the efficiency losses when individual markets face demand and supply shocks but wages are not permitted to play their allocative role. A key element in Freeman and Gibbons' model is that firms and local unions have private information about such shocks. The central union federation then decides on its policy concerning local wage drift anticipating what local bargaining pairs will do, an assumption that gives the model its game theoretic flavor. This model is then used to explain the decentralization of bargaining in Sweden by arguing that local shocks have become more variable as new unions of white collar workers have become more prominent and that the threat of inflation has waned. Thus, the costs of inflexibility have risen as the workforce has become more diverse, while the gains to centralization via a reduction in the threat of inflation have diminished.

It should be noted that the central premise of much of this literature that any intervention which forces higher overall wage levels will lower employment has not gone unchallenged in recent years. Specifically, economists have long recognized that, under conditions of employer monopsony, forcing a company to pay higher wages can lead to higher employment, since such a policy may lower the marginal cost of labor.[6] Card and Krueger (1995) and Manning (1996) argue that the feature of monopsony responsible for such a result – that the firm faces an upward sloping labor supply schedule – characterizes any firm that must expend resources in order to recruit labor. Therefore, it is not necessary for labor markets to be controlled by one employer, an extreme assumption for urban labor markets, in order for elements of monopsony to be present. The upshot of this approach is that interventions forcing higher wages, such as minimum wage laws or anti-discrimination policies directed at raising the relative pay of women or minorities, need not result in employment losses. We return to this point in our consideration of the empirical evidence on the impact of such policies below.

4.3. Employment protection, employment and wages

A substantial body of research attempts to estimate the effect of employment protection on employment and wage levels. A major motivating force behind this work is a desire to ascertain whether the policies of imposing high firing costs and lengthy advance notice on employers, which characterize most OECD countries, have contributed to persistently

[6] Of course, the presence of monopsony does not guarantee that a higher legislated or bargained wage will raise employment. If wages are raised by a large enough increment, the marginal cost of labor will rise and the monopsonist will cut back on the quantity of labor demanded.

high European unemployment rates. This research posits direct and indirect mechanisms by which protection can affect unemployment. The direct effects operate via firms' incentives to hire and fire workers, all else equal, while the indirect effects are due to the impact of firing costs on union wage bargaining behavior.

Regarding the direct effects of higher firing costs, an interesting insight offered by Lazear (1990) is that, in principle, these can be completely offset by the establishment of an appropriate entry fee charged by firms to newly employed workers or, equivalently, by lower starting wages. If firms are able to follow such a policy, the allocation of labor under a system with mandated severance pay is the same as in one without severance pay. However, there may be constraints on the firms' ability to charge an entry fee (or offer a reduced starting wage). For example, workers may be liquidity constrained or worker trust of the firm may be incomplete.[7] In that case, mandated severance pay can have allocative effects. Specifically, we expect severance pay to reduce both layoffs during recessions and new hiring during expansions. Thus, firing costs are unambiguously expected to lower fluctuations in the quantity of labor demanded over the business cycle.

With respect to overall employment levels (the issue of primary interest to those concerned with persistently high European unemployment), a "first order" approach would suggest that, if entry fees or lowered starting wages do not completely compensate for mandated firing costs, then total labor costs will have risen, and we would thus expect lower employment levels (Hamermesh, 1993). However, theoretical analyses of the impact of firing costs on the average quantity of labor demanded suggest that this effect is theoretically ambiguous and will depend on several factors (Lazear, 1990; Bertola, 1992). In particular, Bertola (1992) suggests that the impact of firing costs at given wages (i.e., assuming no offset in pay induced by mandated firing costs) depends on the shape of the marginal product of labor (MRPL) curve and on the presence of discounting and voluntary turnover.

To see this, recall that firing costs will deter both hiring and firing; thus, their net effect on average employment will depend on the relative size of their impact on hiring and firing. Let's begin with the simplest case Bertola analyzes and assume no discounting or turnover. Suppose further that the MRPL curve is relatively steep in the lower employment regions (i.e., during a recession) but relatively flat at high employment levels (i.e., during a boom). Then the number of layoffs deterred by high firing costs will of necessity be relatively small because it would not have taken many layoffs to re-establish equality between marginal productivity and wages in the absence of firing costs. However, the number of new job offers deterred by firing costs will be large because it would have taken many new workers hired to re-establish the equality between marginal productivity and wages. In this example, firing costs will lower the average level of employment. Conver-

[7] A study by Friesen (1996) of the impact of Canadian advance notice and severance pay mandates suggests that the assumption of less-than-complete adjustment of wages is realistic. Friesen finds that such protection lowers the starting wages of non-union, but not of union workers, suggesting that, at least for union workers, starting pay does not fully adjust in response to job protection. Of course, the negative effects on the initial wages of non-union workers may or may not be fully offsetting with respect to the costs of protection.

sely, if the MRPL curve is sufficiently flat during recessions and steep during booms, then firing costs will raise average employment.

Considerations of discounting and voluntary turnover raise the positive effects or reduce the negative effects of mandated firing costs on average employment levels. First, with discounting, the negative effect of firing costs on discharges is increased relative to the negative effect on hiring. This is the case because firing costs must be paid immediately when workers are discharged, while the deterrence to hiring is related to future firing costs. Second, with voluntary turnover, there is some probability that current hires will not need to be discharged; this also raises the magnitude of the negative effect of firing costs on discharges relative to their effect on hiring. If these two effects are large enough, firing costs can actually raise average employment levels even when the shape of MRPL curve alone would not indicate such an outcome.

The second route through which firing costs can affect employment is their indirect effects on wage setting. Lindbeck and Snower (1986), for example, argue that the bargaining position of insiders is enhanced by higher firing costs. They can thus extract more rents, and, in a monopoly union model, this will tend to lower future employment.

4.4. The relationship between wage-setting, centralization and social policies

While bargaining institutions and social policies such as employment protection may each have independent effects on employment and wages, distinguishing these effects in a cross section of OECD countries may be difficult. This is the case because countries with more highly centralized union-management negotiations also tend to have more extensive welfare states with generous social benefits funded by payroll or income taxes. So, for example, wage compression at the bottom of the pay scale could be due to generous UI benefits rather than centrally-determined wage minima. Deciding which factor is more important can be problematic. However, Summers et al. (1993) have devised a model in which wage centralization leads to higher taxes and social benefits. This suggests that the primary causation may indeed flow from institutions that equalize wages to higher welfare state expenditures rather than vice versa.

According to Summers, Gruber and Vergara, this mechanism works as follows. In corporatist societies, such as Sweden, wage setting and labor allocation decisions are determined by groups rather than individuals.[8] Labor's representatives will be less averse to high taxes than individuals because unions recognize the link between taxes and welfare state benefits. Thus, where labor supply levels are set by individuals, as in the United States, the negative effect of higher taxes on labor supply is likely to be greater than in a corporatist society. As evidence for this view, the authors point out that Sweden did not embark on a policy of truly centralized bargaining until roughly 1956. At that time, taxes were only a slightly higher fraction of GNP than in the US. However, after 1956, Sweden's

[8] By "corporatism" is meant highly centralized, coordinated labor-management relations where central labor and employer organizations have considerable authority to impose contract terms throughout the country.

tax rates took off relative to those in the US. Since the Social Democrats were in power both before and after the shift in wage-setting regimes, a change in the governing party cannot account for the increase in tax rates.

If the mechanism outlined by Summers, Gruber and Vergara is correct, then wage centralization would be a fundamental cause of higher levels of social benefits that themselves may feed back and affect the wage distribution. However, a weakness of this argument is that, even in countries such as Sweden, it may be difficult for union federations to control the labor supply of individual members. Moreover, this reasoning is counter to that of Calmfors and Driffill (1988) discussed above that it is precisely the greater sensitivity of encompassing unions to the negative effects on their members of higher taxes (and prices) due to higher negotiated wage bargains that leads them to practice greater wage restraint. Nonetheless, the take-off of social spending in Sweden after the centralization of wage bargaining is intriguing evidence in support of a causal ordering. Perhaps a plausible alternative explanation to that offered by Summers, Gruber and Vergara is that encompassing unions support generous social benefits to deal with the disemployment which results from high wage floors and which would otherwise generate pressure to lower these floors.[9]

5. Wage centralization and macroeconomic performance

As we saw in our discussion in Section 4.2, the opposing effects of bargaining centralization on wage restraint may imply an inverted U-shaped relationship between centralization and unemployment: we expect wages to be most restrained and hence unemployment rates to be lowest in decentralized and in completely centralized systems. Research on this issue has taken the form of small scale international comparisons of up to roughly 20 countries, sometimes in a regression framework. A problem in this literature is the difficulty in operationalizing the concept of centralization. At the two extremes, it is clear that the bargaining regimes of the Scandinavian countries involve more coordination across firms, sectors and unions than the US system of firm or plant bargaining units in the context of a predominantly non-union labor market. However, it may be problematic to decide on a ranking for what scholars tend to believe are the intermediate countries such as Australia, Germany, France or the Netherlands (Calmfors and Driffill, 1988; Soskice, 1990). Yet Calmfors and Driffill (1988) have devised such a ranking based on the degree of cooperation among unions and among employers in wage bargaining.

[9] An additional rationale for a causal link from corporatism to high taxes is provided by Persson (1995). He posits a relative consumption utility model in which individuals receive disutility when others' consumption of goods and services apart from leisure rises. In such a world, by reducing the labor supply of others and therefore their non-leisure consumption, higher taxes have a positive welfare effect that is absent in models where utility depends only on absolute consumption. Persson shows that the more equally distributed are before tax incomes, the larger is the group whose utility will be raised by higher taxes. Thus, if centralization lowers inequality, it will also lower the public's resistance to higher taxes. Of course, this framework depends on individuals feeling envy only about material consumption by others and not about their consumption of leisure.

Several authors have used this approach to measuring centralization in comparative research analyzing macroeconomic outcomes such as the unemployment rate. The long-run differences across countries in unemployment rates identified by cross-sectional regressions may be viewed in the context of "natural rate" theories. In effect, we are attempting to determine whether labor market institutions affect the unemployment rate associated with an economy's macroeconomic steady state. In simple regressions involving under 20 countries, Calmfors and Driffill (1988) and Rowthorn (1992) both found that unemployment did indeed have the expected inverted U-shaped relationship with centralization in the 1980s. However, the estimated shape of this relationship has been found to be very sensitive to how certain countries are classified with respect to centralization; this is a troubling weakness in this body of research. For example, Soskice (1990) makes the case that Japan and Switzerland are not examples of decentralized bargaining as claimed by Calmfors and Driffill (1988), but rather that a high degree of employer coordination in wage-setting in those countries makes them very centralized. When he makes this assumption, the relationship between centralization and unemployment becomes monotonically negative. Whatever the merits of the argument about whether or not Japan and Switzerland have centralizing institutions, Soskice's exercise shows that the results of studies in this area can be extremely sensitive to classification errors.

Other work using this approach also tends to find either a negative or an inverted U-shaped relationship between centralization and unemployment (Calmfors, 1993). But the estimated relationship is also sensitive to whether one separately distinguishes between employer and union coordination (Layard et al., 1991) or between decentralized collective bargaining and non-union wage setting (Nickell, 1997). For example, Layard et al. (1991) find that, all else equal, employer coordination leads to greater restraint (i.e., lower unemployment) than union coordination; and Nickell (1997) finds that union density is positively associated with unemployment, other things equal.

Some of the comparative work on macroeconomic outcomes has explicitly examined the process through which centralization appears to influence unemployment. One mechanism is the rigidity of real wages in the face of macro-shocks such as the oil price increases of the 1970s and early 1980s. Layard et al. (1991) find that, in the 1970s and 1980s, centralized wage setting institutions were associated with more real wage flexibility with respect to the unemployment rate. This likely reflects the fact that wage negotiators at the national level take into account the negative macroeconomic effects of keeping real wages too high in the face of negative demand shocks. And unemployment rose faster in countries with rigid wages. On the other hand, Heylen (1993) found evidence of the inverted U shape in this relationship, in contrast to Layard et al.'s (1991) monotonic positive effect. Again, the sensitivity of the basic results in the face of such a small number of observations is evident.

A final insight into the impact of wage centralization on macroeconomic performance is provided by Freeman (1994). While he did not perform an explicit econometric test, Freeman notes that during the 1980s and 1990s, American real wages have not increased nearly as rapidly as those in Europe, at the same time that US relative unemployment has

fallen (see Table 1). This juxtaposition at least raises the possibility that weaker union power in the US has contributed to its poorer real wage performance (relative to productivity) and that, given negatively sloped demand curves, this has improved the macroeconomic performance of the US relative to other OECD countries. Similarly, in a recent review of the evidence, Blank (1997) gives qualified support for the hypothesis that higher rates of unemployment in Europe and higher wage inequality in the US represent responses to the same fundamental underlying forces – such as technology and trade – conditioned by the degree of flexibility in each labor market.

While macroeconomic performance is one of the more important economic issues one can study through these types of international comparisons, the small number of observations and the presence of confounding variables reduce the robustness of the findings in this body of research. Fortunately, as indicated by our theoretical discussion, there are many other extremely important effects of labor market institutions. And these have been found to be more amenable to focused hypothesis testing with data and methods that allow one to rule out many alternative explanations. We now turn to these other issues.

6. Wage-setting institutions and wage inequality

The most extensive area of research into the labor market impact of institutions concerns their effects on wage inequality. This includes considerations of the effects of both collective bargaining and direct government intervention into wage setting. Research that examines the impact of collective bargaining does not merely compare outcomes for union and non-union workers; it also examines the impact of alternative types of collective bargaining regimes on overall wage inequality, with centralization again being a crucial dimension. Forms of direct government intervention include minimum wage laws as well as anti-discrimination efforts on behalf of specific groups like women or minorities. Moreover, the indirect effects of policies such as active labor market programs on wage inequality have been studied as well and are also considered below, particularly when we discuss responses to labor market institutions in Section 7.

Two types of questions are typically addressed in this literature. First, what is the role of labor market institutions in explaining differences across countries in wage inequality and related outcomes? This is a "levels" question and is thus the most fundamental international comparative question one can ask. This is the strength of this approach. To answer this question, researchers typically try to relate observed differences across countries in the extent of wage inequality and other outcomes to measurable differences in institutions in a manner analogous to studies of the relationship between corporatism and macroeconomic performance discussed earlier. The weakness of this approach is that many things besides the institutions in question may differ across countries, so we cannot be certain if the institutions are really responsible for the observed differences in outcomes. Second, why have wage inequality and other outcomes been changing at different rates across countries, particularly in the 1980s and 1990s? Why, within countries, have some groups like less

skilled workers or women fared differently in recent years from other groups? This approach yields answers to these important policy questions but does not address the more fundamental issues of long-term differences between countries. However, it does have a scientific advantage over the purely cross-sectional approach to the extent that we have noticeable, abrupt changes in policy regimes: in relating such changes to outcome measures, we may be fairly confident that other factors really are held constant (as long as we pay attention to the reasons why the institutional change happened in the first place). The comparative approach can be useful here if one country changes its institutions at a particular time while others do not; we may then apply "differences in differences" techniques to compare the outcome change in the "treatment" country with the outcome change in the "control" country.

6.1. Collective bargaining institutions and overall wage inequality

6.1.1. Overview

Data from the OECD on union density and collective bargaining coverage for 1994 are presented in Table 2. While there is considerable variation across countries, the US ranks very low on both measures, but especially on collective bargaining coverage. These differences suggest that collective bargaining institutions may be an important factor in explaining international differences in wage inequality, especially the considerably higher levels of inequality in the US. There are a number of routes through which this may occur. These are conveniently summarized by the following decomposition of a country's log wage variance:

$$v_i = \alpha_{ui} v_{ui} + (1 - \alpha_{ui}) v_{ni} + \alpha_{ui}(\bar{w}_{ui} - \bar{w}_i)^2 + (1 - \alpha_{ui})(\bar{w}_{ni} - \bar{w}_i)^2 \qquad (1)$$

where for country i, v is the overall variance of log wages; α_u is the fraction of workers who are unionized; v_u and v_n are the variance of log union and non-union wages; $\bar{w}_u$ and $\bar{w}_n$ are average log union and non-union wages; and $\bar{w}_i$ is the country's average log wage level.[10]

By the accounting scheme in Eq. (1), there are several routes through which the industrial relations system can affect overall wage inequality. First, unions typically raise their members' relative wages. This effect alone could increase or decrease overall wage dispersion, depending on where union workers would have ranked in the wage distribution in the absence of unionism. However, as suggested by the final two terms in Eq. (1), in an accounting sense, it is the union–non-union wage gap (not controlling for other wage-influencing factors) itself that is important in "explaining" the overall variance. All else equal, the larger this gap, however it is achieved, the larger the country's overall log wage variance will be.

Second, unions typically negotiate contracts that allow for less variation in pay than occurs in the non-union sector. Freeman (1982) has shown this by examining establish-

[10] This accounting was first used by Freeman (1980) in his analysis of the effects of unions on wage dispersion in the US.

Table 2
Union density and bargaining coverage rates in OECD countries, 1994[a]

	Union density rate	Bargaining coverage rate
Austria	43	98[2]
Belgium	53	90[2]
Denmark	76	90[1]
Finland	81[3]	95[1]
France	9[3]	95[2]
Germany	30	92[1]
Italy	39[1]	82[2]
Netherlands	26[1]	81[1]
Norway	58[1]	74[1]
Portugal	32[1]	50[4]
Spain	22	66[1]
Sweden	91	93[1]
Switzerland	26	50[2]
UK	36	47[2]
Australia	35	80
Canada	38[1]	36
Japan	24[2]	22
New Zealand	31	31[1]
Non-US average (unweighted)	42	71
US	16	18

[a] Source: Blanchflower (1996), based on OECD data. Notes: 1, 1993; 2, 1992; 3, 1995; 4, 1990.

ment data for the US in the 1970s. Similar results were obtained for the 1980s in the UK (Gosling and Machin, 1995) and in Italy (Dell'Aringa and Lucifora, 1994). More broadly, using microdata on individual union and non-union workers, Blau and Kahn (1996b), found this to be the general pattern in the 1980s within a number of countries, including the US, UK, Austria, Switzerland, West Germany, Hungary and Norway. Therefore, even if the variance of the log of both union and non-union wages were the same in two countries, greater union density would lead to a smaller overall variance: where unions are more prevalent, the lower union variance in pay would get a larger weight in Eq. (1). However, the dispersion of wages within each sector can in fact differ across countries. Such differences in within sector variances constitute a third route whereby wage-setting institutions can influence overall wage inequality. The extent of wage dispersion within both the union and non-union sectors is in turn likely to be affected by the practices discussed above concerning centralization and contract extension to non-union workers.

First, with respect to the union sector, coordination across bargaining units is expected

to reduce union wage inequality by lowering interfirm and interindustry wage differentials. And, a substantial portion of the wage inequality we observe in the US, for example, is associated with such firm or industry wage effects (Blau, 1977; Krueger and Summers, 1988; Davis and Haltiwanger, 1991; Groshen, 1991). In this regard, the US is likely to constitute one extreme since collective bargaining in the US is relatively decentralized, with an emphasis on single-firm agreements which, in most cases, are not firm-wide (Hendricks and Kahn, 1982). In contrast, in most of the other OECD countries, bargaining is generally conducted at least on an industry-wide level with the economy-wide bargaining in the Scandinavian countries constituting the other extreme. In such countries industry- or economy-wide collective bargaining agreements are signed by a union or a union federation and an employer federation.[11] The agreements typically call for wage minima, and in several countries cover virtually everyone in the industry regardless of union membership. As noted earlier, Table 2, for instance, shows collective bargaining coverage at or above 90% in many OECD countries. In addition, in some cases, the sectoral agreements stipulate wage scales beyond the minimum, and individual company-level contracts often supplement the basic industry-wide contract. The question of whether these industry-wide or economy-wide wage minima actually affect the wage distribution or are undone by company-level bargaining or negotiated increases for higher-paid workers can only be resolved empirically.

If the negotiated wage minima, which apply across diverse units, are actually binding, they will tend to disproportionately bring up the floor among workers covered by the contract. Thus, while Eq. (1) refers to the variance of wages, consideration of wage floors leads us to expect a greater narrowing at the bottom than at the top in the union sector. Explicit attempts by union movements and governments in several countries to assist low wage workers will further reinforce this effect. Similarly, other institutions, such as UI or active labor market policies, may place a floor under wages for the purposes of wage bargaining, with possibly large relative effects for workers at the bottom of the wage distribution. This suggests that we need to examine the entire wage distribution rather than focusing solely on the variance in assessing the impact of institutions. Quantile regression techniques such as those used by Chamberlain (1991), Blau and Kahn (1996b), and Kahn (1998a,b), or full-distributional accounting methods such as those devised by Juhn et al. (1991), DiNardo et al. (1996), and DiNardo and Lemieux (1997) become essential in studying characteristics of wage distributions beyond their variances.

Second, with respect to the non-union sector, collective bargaining arrangements can have important impacts on the variance of non-union wages. Again, the US is likely to constitute one extreme where, given the small size of the union sector and the lack of any formal arrangements to extend contracts to non-union workers, the impact of unions on the variance of non-union wages is likely to be small. However, in many OECD countries, including Austria, Belgium, Germany, Italy, the Netherlands and Switzerland, among

[11] These sectoral agreements are typical in Austria, Belgium, France, Germany, Italy, Scandinavia and several other OECD countries. For a more detailed discussion of specific collective bargaining mechanisms, see Katz (1993), the chapters in Freeman and Katz (1995), and EIRR (October, 1992).

others, the government routinely extends the terms of collective bargaining agreements to non-union workers (EIRR, 1992; Wallerstein et al., 1997). Table 2, for example, shows the very large differences that sometimes are observed between union membership and union coverage, some of which reflect formal contract extensions. To the extent that unions in all countries tend to compress wages at the bottom in the union sector, contract extension will not only reduce wage variation in the non-union sector, but compress wages at the bottom as well. In addition, there is some evidence that non-union firms tend to imitate union wage structures as well as union wage levels. For example, for the US, Kahn and Curme (1987) found that, other things equal, the larger an industry's union density, the smaller the standard deviation of the log wages of that industry's non-union workers; they instrumented for the industry's union density in obtaining this result. In general, we would expect that the higher union density in other OECD countries should produce considerably more "voluntary" imitation of union pay structures by non-union firms in those countries than in the US, even in the absence of formal contract extension. Moreover, as noted above, government policies may further reinforce the resulting tendency towards wage compression at the bottom.

A final point to note with respect to the impact of international differences in industrial relations systems on wage inequality is that each of the three components identified in Eq. (1) – the union–non-union differential, union density, and variance of wages within the union and non-union sectors – may be conceptualized as having two components: one attributable to cross-country differences in worker characteristics and the other due to international differences in behavior or prices. Since it is the latter that more accurately reflect the fundamental causal effect of unions, we will focus on them for the most part in our discussion of the evidence below.

Before turning to a detailed consideration of each of these routes by which union wage-setting institutions may influence wage inequality, we provide an overview of some of the international evidence, focusing primarily on the difference between the US and other OECD countries. As our preceding discussion suggests, this is an especially instructive comparison because the US constitutes an extreme case of highly decentralized wage setting. We draw on results presented in Blau and Kahn (1996b) to address two questions. First, are the general patterns of differences in wage inequality between the US and other OECD countries consistent with the role of institutions sketched above? Second, in terms of the three routes by which differences in wage setting institutions can influence differences in inequality, how important is each, in an accounting sense, in explaining the differences in wage inequality between the US and other countries?

Fig. 1 shows a number of measures of male log wage inequality for the US and a number of OECD countries and Hungary. The figure indicates that the US has a considerably higher level of overall wage inequality than the other countries. Panels (a) and (b) of Fig. 1 show that both the standard deviation of log wages and the 90–10 percentile log wage differential are considerably greater in the US than in the other countries. Significantly, however, this higher level of inequality reflects considerably more compression at the bottom of the distribution in the other countries relative to the US, but a much smaller

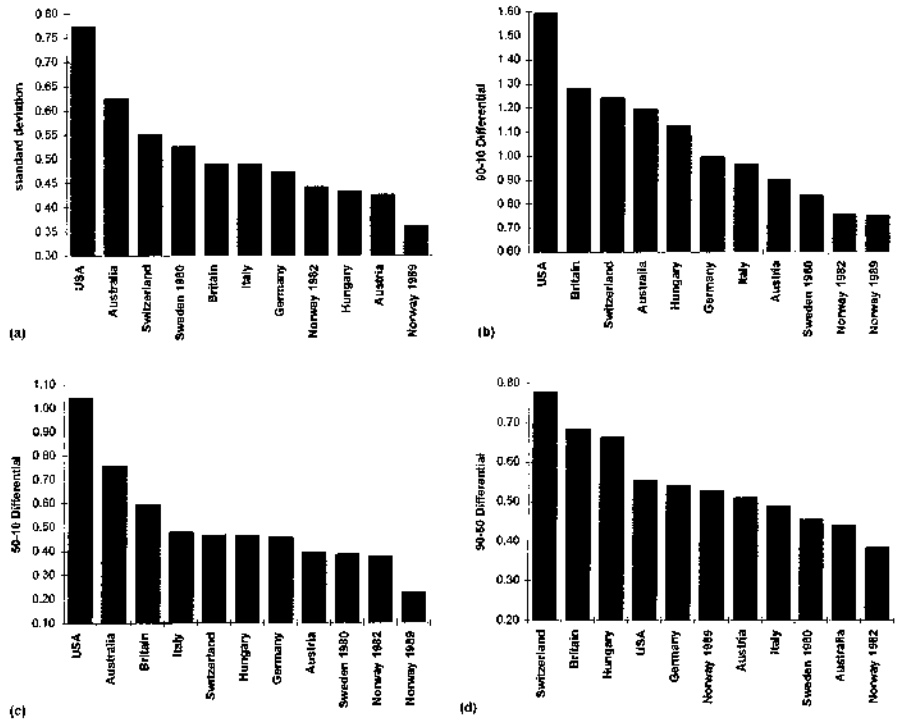

Fig. 1. Summary measures of male log wagw inequality in OECD countries. Source: Blau and Kahn (1996b). Except as indicated, data are from the mid to late 1980s. For details of the construction of the wage measures, see Blau and Kahn (1996b).

difference in the degree of wage inequality at the top of the distribution. Thus, while the 50–10 percentile wage differential is much larger in the US than elsewhere (panel c), the US 90–50 differential is quite similar to that in several of the other countries (panel d). This pattern is quite consistent with what we would expect based on the highly decentralized wage-setting institutions in the US.[12] Moreover, Blau and Kahn (1996b) found that cross-country differences in the distribution of measured human capital characteristics of workers, as well as their distribution across industries and occupations, could account for only a small portion (under 1%) of the pattern of higher dispersion of wages and a larger but minority share (35–43%) of the wider spread between the middle and the bottom of the wage distribution in the US than elsewhere.

[12] Published data show a similar pattern. Specifically, the tabulations in the OECD (1993) indicate that the 50–10 gap was far larger in the US than in other OECD countries. While the 90–50 gap was also larger in the US, the difference between the US 90–50 gap and that of other countries was much smaller than the corresponding difference for the 50–10 gap.

There is also some general evidence that the degree of centralization of wage-setting institutions tends to be associated with lower overall wage inequality, and that this effect particularly reflects a greater compression at the bottom of the wage distribution. For the countries shown in Fig. 1, Blau and Kahn (1996b) found that centralization was significantly negatively related to overall male wage inequality and to the pay differential between men at the 50th and at the 10th percentiles of the wage distribution, as well as to the portion of the 50–10 gap not accounted for by differences in the distribution of measured personal characteristics. They used as their measure of centralization an unweighted average of the rankings of a number of industrial relations researchers presented in Calmfors and Driffill (1988, p. 18), correcting for differences across authors in the number of countries ranked. Similarly, for a sample of 17 countries, Rowthorn (1992) found a significant negative correlation between centralization and interindustry wage dispersion using Calmfors and Driffill's (1988) ranking as his measure of centralization. Finally, similar results were obtained by Bell et al. (1996) in their comparison of UK and France. With France's greater coverage by collective bargaining and higher minimum wages, its labor market prices served to lower the dispersion of its wage distribution relative to that of UK in 1986 and 1992. In particular, they found that the bottom of the British wage distribution would have been considerably raised with French wage-setting institutions.

In terms of the relative importance of the various ways in which institutional factors may influence international differences in inequality, Blau and Kahn (1996b) found that the most important factor was the higher variance in wages within the union and non-union sectors in the US compared to the other countries, accounting for 86% of the US-other country wage dispersion difference, on average.[13] Controlling for the distribution of human capital characteristics and occupational and industrial location did not alter the pattern of higher wage variances within each sector in the US. Consistent with our expectations based on the impact of unions, the difference between the US and other countries was largest for those at the bottom of the distribution. Further, the differences between the US and other countries were found to be more pronounced for non-union than for union workers, implying greater spillover of union wage structures to the non-union sector in the European countries than in the US.[14] With respect to the other factors identified in Eq. (1), Blau and Kahn (1996b) found that lower union density in the US than elsewhere accounted for 12% of the US-other country difference, on average, while the higher US union–non-union differential accounted for less than 2%. We now turn to a more detailed consideration of each of the factors identified in Eq. (1).

6.1.2. The union–non-union wage differential

Looking first at the union–non-union wage differential, there is considerable evidence that

[13] The countries included in this portion of their analysis were the US, UK, Austria, Switzerland, West Germany, Hungary and Norway.

[14] See also Kahn (1998b).

the ceteris paribus gap varies substantially across countries. Two comprehensive studies of union–non-union wage differentials across countries based on microdata concluded that the US had the largest union–non-union differential, other things equal, followed by UK. Blanchflower and Freeman (1992) found that, controlling for human capital characteristics, the union–non-union wage gap was about 20% in the US for the 1985–1987 period, compared to 10% in the UK, and 4–8% in Austria, Australia, Switzerland, and West Germany. Blanchflower (1996) updated this study using data through 1993 and a larger number countries. The results again showed larger differentials in the US and the UK, which had ceteris paribus union–non-union pay gaps of about 15% and 10%, respectively.[15] These two studies are particularly noteworthy because they were conducted on microdata sets that were compiled for the explicit purpose of making international comparisons and thus were as comparable as possible. Further, single country studies of the impact of unions on relative wages have largely found similar results, with the exception of findings for Canada, which have tended to be in the 10–20% range – larger than the 5% effect estimated by Blanchflower (1996).[16]

Eq. (1) suggests that the higher estimated union–non-union wage gaps in the US, the UK and Canada serve to raise their levels of overall wage dispersion relative to those in other OECD countries. Moreover, since these results are based on membership, not collective bargaining coverage, relatively small union–non-union wage differentials are an expected outcome in countries with contract extensions to non-union workers or with very high levels of union coverage which would especially prompt imitation of union wage setting by the non-union sector. While there is considerable evidence of a higher union wage premium in the US, as we have seen, only a negligible portion of the higher US wage dispersion can be explained by this factor (Blau and Kahn, 1996b). We now turn to a consideration of the other components of the decomposition in Eq. (1).

6.1.3. Union density and collective bargaining coverage

Looking at the data in Table 2, we again note the especially low ranking of the US. Only 16% of US workers were union members and just 18% were covered by collective bargaining contracts compared to an average union density of 42% and collective bargaining coverage rate of 71% for the other countries. This difference reflects the presence in other countries of contract extension and other mechanisms by which the union agreement is binding on non-union workers. There were declines in union density in the 1980s in a number of countries, including the US where unionism fell by 6 percentage points. Decreases in collective bargaining coverage occurred as well among a smaller group,

[15] The countries included Australia, Austria, Canada, Ireland, Israel, Italy, Japan, the Netherlands, New Zealand, Norway, Spain, Switzerland, the UK and the US. Ireland and Japan had larger union wage gaps than the US, but Blanchflower (1996) considers these results to be aberrations caused by inadequate controls for industry and firm size.

[16] See for example, Main (1996) for the UK; Schmidt (1995) or Schmidt and Zimmermann (1991) for Germany; Lemieux (1993), Green (1991), or Simpson (1991) for Canada; Kornfeld (1993), Christie (1992), Miller and Mulvey (1993), or Mulvey (1986) for Australia; and Dell'Aringa and Lucifora (1994) for Italy.

including the US where coverage fell by 8 percentage points and the UK where there was an especially large drop of more than 20 percentage points.

In addition to the US, the UK and Canada had among the lowest rates of union density and collective bargaining coverage of the countries listed. The combination of low union density with high union wage premiums in these three countries raises the possibility that their large estimated union effects are due to selectivity, i.e., only the strongest unions survive pushing up the estimated return. An earlier empirical literature on the selectivity issue was reviewed by Lewis (1986) who concluded that unions in the US still appear to have a significant positive relative wage effect of at least 10%. More recently, Blanchflower (1996) has pointed out that union coverage in the 1980s fell the most in countries with relatively high union–non-union wage differentials – the US, the UK and Austria. This pattern raises the possibility that some of the negative cross-sectional relationship between the ceteris paribus union–non-union pay gap and union density was caused by union firms moving up their labor demand curves in response to high union relative wages.

In terms of the quantitative importance of differences in union density in explaining international differences in wage dispersion, as we have seen, the large US-other country differences in union density explained relatively little (12%) of the higher US log wage variance compared to other countries (Blau and Kahn, 1996b). On the other hand, Lemieux (1993) found that US-Canadian differences in union coverage accounted for 40% of the difference in wage dispersion between the two countries in the 1980s. It is likely that the differences between these two studies reflect the countries on which they focus. Within sector wage distributions are likely more similar for the US and Canada than is the case for the US relative to the more heterogeneous countries included in the Blau and Kahn study,[17] while Canadian workers were about twice as likely to be union members as American workers.

Another take on the impact of union coverage on the dispersion of wages is provided by an examination of changes within a country over time. For example, in Australia, the US and the UK, union density fell while wage inequality rose in the 1980s, with especially large changes in the latter two countries. Several studies have sought to determine the role played by the decline in union coverage in explaining rising wage dispersion within these countries. In effect, these studies simulate what would have happened to overall wage inequality if union density had not declined, given the actual changes in the distributions of wages within the union and non-union sectors.

Looking first at the US, using an accounting scheme like that shown in Eq. (1), Freeman (1993) and Card (1996) both found that declining unionization over the 1980s accounted for 20% of the rise in wage inequality. Somewhat smaller impacts were obtained for the US by DiNardo et al. (1996) who departed from this accounting framework and created entire simulated distributions of log wages under the assumption that unionization had not changed between 1979 and 1988. They found that deunionization in the US contributed to about 3% of the rise in the standard deviation of log wages for women and 14% for men.

[17] Again, these included the US, UK, Austria, Switzerland, West Germany, Hungary and Norway.

Estimates in this range (13–21%) are obtained for UK by Schmitt (1995) who used an analysis similar in spirit to Eq. (1) to examine the impact of falling union density in UK from 1978–1980 to 1986–1988 on wage differentials by education and occupation. Finally, a somewhat higher estimate is obtained for Australia by Borland (1996), also using this accounting framework; he found that declining unionization in Australia accounted for 30% of the observed increase in log wage dispersion over the 1986–1994 period. However, it should be noted that, while Borland's estimates of the share of the change in dispersion due to declining union density are larger than those obtained for the US and UK, the declines in union density and the increases in inequality in Australia were much smaller.

A comparison of the US and Canadian experience over the 1980s is particularly instructive because the economic and labor force structures of the two countries are similar and they are major trading partners, yet wage inequality increased by considerably more in the US than in Canada. For example, between 1979 and 1987, the variance of the log of male full-time, full-year earnings rose by 0.034 in the US compared to only 0.018 in Canada (Blackburn and Bloom, 1993, p. 254). And, during the 1980s, US union density fell relative to that in Canada. Using the framework in Eq. (1), Lemieux (1993) finds that this relative fall was responsible for 40–45% of the growth in wage inequality in the US relative to Canada.[18] This closely matches the magnitude of his estimate of the impact of union density in explaining the difference in wage dispersion between the US and Canada at a point in time discussed above.

Each of the country studies reviewed above which examined changes over time attributed a substantial portion of a country's increase in wage inequality in the 1980s simply to the fact that unionization declined, while a similar approach suggests that a larger decrease in union density in the US explained a sizable share of that country's rise in wage inequality compared to Canada. However, it is likely that the decline in unionization itself was caused in part by the impact of international trade and technological change, both of which reduced the demand for blue collar manufacturing workers, a group that is traditionally relatively highly unionized. It is thus possible that the results obtained for the effect of deunionization are upper bounds for the true effects of unionization. In this regard, some recent results for the Canada-US comparison provide some reassurance. Riddell (1993) finds that the higher degree of unionism in Canada than in the US is primarily due to the more favorable legal environment there, rather than to differences in the structure of the economy. For this case at least, we have some confidence that there really is an independent effect of unionization.

While the possible endogeneity of changes in unionization may lead to an upward

[18] In a related study, DiNardo and Lemieux (1997) attempted to explain Canada's slower growth in male log wage inequality than that of the US over the 1981–1988 period. Using a methodology similar to that employed by DiNardo et al. (1996), they attributed about one third of Canada's slower increase in the log wage variance of males to the combined effects of its greater unionization rate and its more equalizing union pay effects; another third was attributed to the declining real value of the minimum wage in the US. Again, institutions were important in explaining the different outcomes in Canada and the United States.

biased estimate of the effects of these changes, there is an opposing factor that implies a downward bias. Specifically, each of the accounting exercises discussed above assumes that unions have no effect on the distribution of non-union wages. However, as we have seen, there is evidence that non-union firms do indeed imitate union wage structures as well as union wage levels. This implies that deunionization will raise the dispersion of non-union wages and also suggests that, in international comparisons, there are more spillovers of the union pay structure to non-union workers in more highly unionized countries, even when there is no explicit government policy to extend union contracts to the non-union sector. For example, Abraham and Houseman (1995) find that in Germany, voluntary imitation is so common that formal extension of union contracts by the government is rarely needed. The spillover of union wage structures to the non-union sector implies that estimates of the impact of deunionization on rising wage inequality or on international differences in wage inequality which take the non-union wage variance as given may be too small.

6.1.4. Wage dispersion within the union and non-union sectors

The remaining portion of the decomposition of a country's wage dispersion implied by Eq. (1) is related to the levels of wage inequality within the union and non-union sectors. As we have seen, comparative research shows this to be, in an accounting sense, the most important reason for the higher level of wage dispersion in the US than in other OECD countries accounting for 86% of the US-other country difference in log wage variance, on average (Blau and Kahn, 1996b). Further, Lemieux (1993) finds, that for the late 1980s, higher within sector variances account for at least 60% of the higher log wage variance in the US relative to Canada. We now turn to a more detailed consideration of the evidence regarding the impact of institutions on the dispersion of wages within the union and non-union sectors.

One of the most important aspects of coordination is that in the countries with the most centralized wage-setting, wages in different industries are set in the same contract. We thus expect less interindustry wage dispersion than under decentralized bargaining structures. Of course, wage drift at the plant level could theoretically undo the effects of frame wage agreements; thus, a comparison of industry wage differentials across countries with different bargaining regimes provides a test of whether the institutions make any difference in practice. And analyses of industry wage differentials in the 1980s do indeed show that they were considerably smaller in corporatist countries such as Sweden, Norway, Austria or Germany than in countries with less centralized bargaining such as the US.[19] It has also been found that the UK, with levels of unionism and coordination between that in the US and other OECD European countries, also had ceteris paribus industry wage differentials between the corporatist countries and the US (Kahn, 1998b).

While these studies indicate that industry wage differentials are smaller where bargain-

[19] See Albaek et al. (1996); Edin and Zetterberg (1992); Barth and Zweimüller (1992); Zweimüller and Barth (1994); and Kahn (1998b).

ing is more coordinated across sectors, we also expect high wage floors in such frame agreements to disproportionately affect workers at the bottom. Consistent with this expectation, Kahn (1998b) found that the dispersion across industries of the 10th percentile of the conditional log wage distribution (i.e., controlling for other factors such as human capital characteristics, union membership, and occupation) was much greater in the US than in other countries such as Germany, Austria, Sweden, Norway, and UK; in contrast US-other country differences in the interindustry dispersion of other quantiles of the conditional log wage distribution such as the 50th or the 90th, as well as its mean, were considerably smaller than that at the 10th percentile. The greater wage coordination in corporatist countries, particularly in Sweden and Norway, was especially evident for those at the bottom of the distribution. And, where data were available, similar patterns were found for non-union as well as union workers, consistent with spillovers and contract extensions (Kahn, 1998b).

Additional evidence on the importance of institutions for within sector wage inequality comes from instances where individual countries changed their wage-setting regimes. We can compare wage inequality before and after such changes; further, these differences can be contrasted with changes in inequality in other countries where such a change in institutions did not occur. A focus on changes over time allows any unmeasured characteristics of a given country which would otherwise affect its wage distribution to be "differenced out," enabling us to concentrate more precisely on the impact of the institutional change in question.

Sweden is a particularly interesting case because it has experienced episodes in which institutional changes occurred which would be expected to decrease inequality, as well as others which would be expected to increase it. While Sweden has had centralized wage setting since at least the 1950s, during the 1964–1983 period, its major blue collar union, the LO, embarked on a "solidarity wage" policy of radical equalization of pay by giving especially large increases to the lowest paid workers. This new wage policy involved equal kroner/hour wage increases instead of percentage increases and special funds to raise the wages of wages low paid workers (Edin and Topel, 1997). Hibbs (1990) shows that the coefficient of variation of blue collar union wages took an abrupt and large downturn precisely in the mid-1960s, following an eight year period of gradually rising dispersion. Further, Edin and Topel (1997) and Edin and Holmlund (1995) document the sharp decline in the returns to education in Sweden following the 1960s, while Edin and Topel (1997) present evidence that interindustry wage differentials contracted sharply between 1960 and 1970. The abrupt nature of the change in bargaining practices and the correspondingly sudden decrease in wage inequality following these changes constitute fairly strong evidence that the institutional change had some impact on the wage distribution. In Section 7.1, we discuss the negative relative employment effects of this change in wage-setting policies for low-paying industries. The employment findings provide further evidence suggesting that the solidarity wage policy did in fact alter the wage distribution.

Perhaps in response to the strains caused by wage leveling, in 1983, the Swedes abandoned the country's economy-wide wage setting practices and moved to a system of

industry-wide wage bargains. In principle, this structure can allow more interindustry wage variation. And the Swedish wage distribution did abruptly become more dispersed following 1983 (Hibbs, 1990; Edin and Holmlund, 1995). Again this pattern is consistent with a real effect of changing bargaining institutions on the wage distribution. However, Edin and Holmlund (1995) show that, in the 1980s, supply and demand were changing to the detriment of low skilled workers and this could be a competing explanation for rising Swedish inequality during this period.

The Swedish experience of the 1980s points up a difficulty in estimating the impact of changes in wage-setting institutions when supply and demand forces go in the same direction as the institutional changes and may in fact have contributed to the institutional changes. Indeed, in the Swedish case, Edin and Topel (1997) note that excess demand for skilled workers in the early 1980s (partly due to wage leveling in encompassing labor agreements) helped lead to the end of economy-wide bargaining in 1983. In contrast, the case of Norway in the 1987–1991 period provides an interesting instance in which institutions changed in the opposite direction to supply and demand forces and opposite to institutions in virtually all other advanced countries (Kahn, 1998a). Until 1982, Norway's collective bargaining system was quite similar to Sweden's in that there were economy-wide centralized negotiations between national union and employer federations. And, like Sweden, as well as several other countries,[20] collective bargaining became less centralized in Norway during the 1980s. Decentralization took the same form as it did in Sweden – industry-wide bargains replaced the economy-wide agreement. However, spurred by the recession brought on by reduced oil prices after 1986, in 1988, the national government in Norway took steps to recentralize the country's bargaining system.[21] In 1988 and 1990, negotiations returned to their nationwide level, and low paid workers received higher absolute (and therefore percentage) wage increases than others did.

Supply and demand for low skilled labor in Norway changed during the late 1980s and early 1990s in ways similar to that in other countries, notably Sweden (Kahn, 1998a). And, at a time when bargaining structures were breaking apart in other countries, Norway's was becoming more monolithic. Consistent with this change, Norway was the only OECD country with a sharply narrowing gap between the middle and the bottom of the wage distribution during the 1987–1991 period. And Kahn (1998a) finds that a fall in the price of skills contributed importantly to this reduction in inequality, as would be expected based on the wage policies adopted by the union federation in this period. Moreover, while the supply and demand for skills in Norway changed similarly in both the 1980–1983 and 1987–1991 periods, in the earlier period, when bargaining was being decentralized, the return to skills rose. These comparisons of Norway with Sweden and for Norway during a period of recentralization with a period of decentralization provide evidence that the change in Norway's bargaining structure did narrow wage differentials. As in the case

[20] These countries include the US, the UK, Italy, West Germany, and Australia (Katz, 1993).

[21] The government's goal here was wage restraint, and recentralizing negotiations with special wage increases for the low paid was deemed necessary in order to get union cooperation in the effort (Kahn, 1998a).

of Sweden, we discuss employment responses to the declining wage differentials caused by changes in bargaining regimes in Section 7.1.

A final example of a wage-setting institution, in this case a government intervention, that appears to have had a narrowing effect on the wage distribution is Italy's system of wage indexation, the scala mobile, which was in place from 1975 to 1992. This was a nationally-mandated cost of living adjustment that explicitly gave low paid workers larger relative increases than others (Erickson and Ichino, 1995). Evidence of the impact of this policy is provided by comparing Italy to other countries. There was rapid inflation from 1975 to 1983 in Italy, averaging 10–20% per year, yet wage inequality fell sharply during this time, in contrast to virtually all other OECD countries (OECD, 1993). Moreover, through 1987, the Italian wage distribution did not widen, in contrast to the US and many other countries, even though supply and demand for skills in Italy changed in qualitatively similar ways to the American experience (Erickson and Ichino, 1995, p. 296). Again, a strong case can be made for asserting the impact of an institutional change on the wage distribution. Analysis of changes in the Italian wage structure after the end of indexation in 1992 would provide further evidence on the importance of institutions and could thus be a very fruitful area for future research.

6.2. *Wage-setting institutions and the relative pay of particular groups: women*

An implication of wage setting mechanisms that bring up the bottom of the wage distribution is that they increase the relative wages of low skill workers. Further, if some workers are confined by employer or union exclusion or other factors to relatively low-paid sectors of the economy, coordinated wage bargaining systems that reduce intersectoral pay differences will raise these workers' relative pay as well. This reasoning has special force for male-female differentials, since, in all countries, women have less labor market experience (an important dimension of skill) than men, on average, and tend to be located in lower paid industries and occupations (e.g., Blau et al., 1998). Thus, while research on the sources of the gender pay gap has traditionally focused on what might be termed "gender-specific" factors, particularly gender differences in human capital characteristics and differences in the treatment of otherwise equally qualified male and female workers (i.e., labor market discrimination), wage structure is an additional and possibly important determinant. Wage structure describes the array of prices set for various labor market skills (measured and unmeasured) and rents received for employment in particular sectors of the economy.[22] This in turn becomes a mechanism whereby labor market institutions can influence the gender pay gap, or indeed any demographic differential.

Given the considerable variation in wage setting institutions across countries, it makes sense to search for this effect in the context of international comparisons, again noting that the position of the US at one extreme with highly decentralized wage-setting institutions

[22] Important work on trends over time in the black-white pay gap by Juhn et al. (1991) was the first to point to the importance of overall wage structure for the relative wages of demographic groups.

Table 3
Gender wage ratios and female percentiles in the US and Sweden, 1984[a]

	United States	Sweden
Female-male log wage ratio		
Unadjusted	66.9	82.2
Adjusted[b]	82.7	90.9
Mean female percentile in:		
Male log wage distribution	29.6	29.9
Male residual distribution[b]	36.6	37.4

[a] Source: Blau and Kahn (1996a).
[b] Based on hourly earnings adjusted for education, actual experience and its square, major industry and occupation.

makes for a natural reference point. And, indeed, in the mid-to late 1980s, Blau and Kahn (1996a) found that the gender pay gap was higher in the US than in most other OECD countries, although American women appeared to have the same or more labor market attachment, were more likely to work full-time, and were no more and frequently less segregated into traditionally female sectors than women other countries.[23] This apparent paradox was explained by the lower prices of labor market skills and rewards to employment in high-paying sectors in the other countries: if the US had other countries' prices of skills and returns to industry, occupation and union status, then the gender pay gap in the US would be as low or lower than that in any of the other countries studied. In conjunction with the international evidence discussed above on the importance of wage-setting institutions in influencing labor market prices, Blau and Kahn's (1996a) finding implies that, in addition to lowering pay differentials along other dimensions, centralizing wage-setting institutions greatly lower the gender pay gap.

Table 3 illustrates these findings for a comparison of the US and Sweden. This comparison is of particular interest because the US and Sweden represent cases at the extremes of an international ranking of both wage centralization and the female-to-male wage ratio, with the US having highly decentralized wage-setting and a high gender gap while the opposite is true for Sweden. The table shows that the gender ratio is 16 percentage points higher in Sweden than in the US, and it remains 9 percentage points higher after adjusting for gender differences in measured characteristics.[24] Insight into the role that wage setting institutions play in producing these large differences in the gender gap may be gained by looking at the percentile rankings of women in each country's male wage distribution. Gender-specific factors – i.e., gender differences in qualifications and the extent of labor

[23] The other countries were Australia, Austria, UK, Hungary, Italy, Norway, Sweden, Switzerland, and West Germany.

[24] For each country, the adjusted wage ratio is $\exp(X_f\beta_f)/\exp(X_f\beta_m)$ where X_f is a vector of means of the explanatory variables for women, β_m and β_f are vectors of estimated coefficients from log wage regressions estimated for men and women separately.

market discrimination – will influence the placement of woman in the male wage distribution, while wage structure will determine the size of the wage penalty associated with this location.

Surprisingly, given the large differences in gender ratios shown in Table 3, the mean percentile of women in the overall male wage distribution[25] and the residual male wage distribution[26] is virtually identical in the two countries. This suggests that the large difference in the gender gaps between them is entirely due to the larger wage penalty placed on women's lower position in the male wage distribution in the US than in Sweden. Fig. 2 further illuminates the impact of wage compression at the bottom of the distribution in producing this outcome. It presents the female cumulative distribution functions that result from placing women in male wage deciles on the basis of male log wage cutoffs. While the US female cumulative distribution function is quite similar to that of Sweden, a larger proportion of women are in the lowest male wage decile in Sweden (29%) than in the US (20%). This suggests that women particularly benefit from formal or de facto wage floors in Sweden that lessen the wage penalty for those at the bottom.

Other analyses of women's relative wages come to largely the same qualitative conclusion regarding the importance of wage-setting institutions. First, in a 1989–1990 comparison of Australia and Canada, Kidd and Shannon (1996) found that the Australian gender pay gap was considerably smaller than that in Canada (by about 0.14 log points). Australia's more compressed wage structure explained 37–66% of this difference. The Australian institution of nationally-binding wage awards issued by government tribunals as well as its higher level of unionization are likely candidates for its lower wage dispersion. Second, Edin (1993) studied the effect of Sweden's solidarity bargaining on the gender pay gap. During the key 1968–1974 period, when collective bargaining agreements were compressing the wage distribution at the bottom, the gender pay gap in Sweden fell by 0.062 log points. Of this decline, 82% was due to the compression of the wage structure, which raised the pay of low wage workers in general, including women. Finally, Hunt (1997) finds that the extension of western Germany's relatively high union wages to eastern Germany with monetary union reduced the gender pay gap by roughly 10 percentage points.

While changes or differences in union wage-setting arrangements have had an important impact on gender pay gaps, in several instances governments have intervened on behalf of women through anti-discrimination policy. It is likely that the nature of wage-setting institutions can have an important influence on the impact of anti-discrimination policies. Specifically, if there are mechanisms through which wages can be centrally altered, then a policy of reducing gender wage differentials can have a more immediate effect than when wage-setting is decentralized. This interaction between gender-specific policies of anti-

[25] For each country, this is obtained by assigning each woman a percentile ranking in that country's male wage distribution and finding the female mean of these percentiles.

[26] For each country, this is obtained by assigning each woman a percentile ranking of her wage residual (from the male wage regression) in the distribution of male wage residuals (from the male wage regression) and finding the female mean of these percentiles.

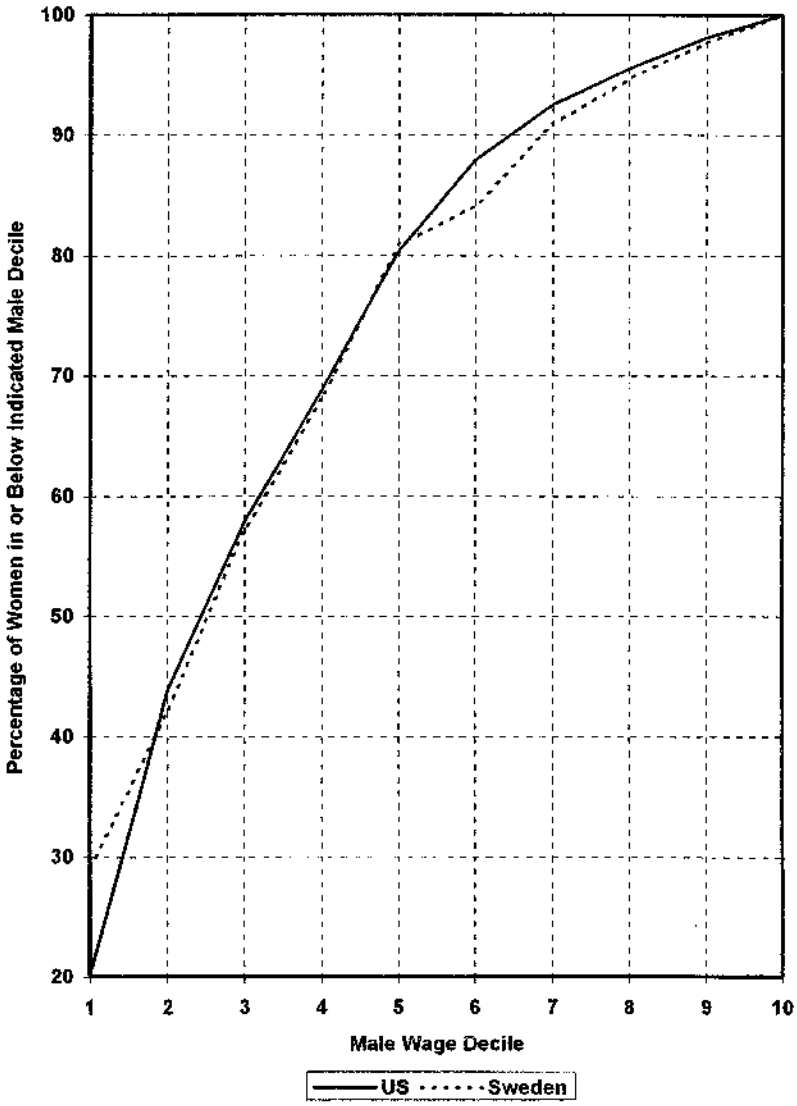

Fig. 2. Cumulative distribution function, female wages relative to the male wage distribution. Source: Blau and Kahn (1996a).

discrimination law and general wage-setting mechanisms may help explain the relative effectiveness of government attempts to reduce the gender pay gap in the US, the UK and Australia.

In the US, the Equal Pay Act of 1963 and the Civil Rights Act of 1964 outlawed pay, employment, layoff, and promotion discrimination on the basis of race, gender or national origin. Significantly, these laws predate those in other advanced countries. Enforcement was to be accomplished through individual or group lawsuits. An additional measure was an Executive Order implemented for women in the early 1970s which added the threat of loss of government contracts for firms that discriminated. There is some econometric evidence that, all else equal, government policy in the 1970s raised the female/male pay ratio (Beller, 1979); and discrimination as conventionally measured declined in the 1970s and 1980s as well (Blau and Beller, 1988; Blau and Kahn, 1997). Moreover, in some local or state government jurisdictions, authorities have mandated comparable worth – equal pay for work of equal value – for government workers, with some positive earnings effects on disproportionately female public sector jobs (Killingsworth, 1990). Yet it was not until the 1980s that the gender pay gap in the US fell. It is of course possible that the removal of discriminatory barriers in higher education, as well as the anti-discrimination legislation generally, indirectly encouraged women to accumulate higher levels of training and other human capital and ultimately contributed to the falling gender gap in the 1980s (Blau and Kahn, 1997; O'Neill and Polachek, 1993). Yet, perhaps because of the individualistic nature of the enforcement of the law, the immediate wage effects of the US legislation appear small. This contrasts with the implementation of anti-discrimination policies in Australia and the UK where the more highly centralized wage setting institutions in these countries appear to have resulted in much larger immediate effects.

The Australian laws mandating equal pay for equal work and equal pay for work of equal value were implemented during the 1969–1975 period. The latter essentially involved implementing a policy of comparable worth on a nation-wide basis. The enforcement mechanism was Australia's system of government tribunal awards which set occupational pay rates estimated to cover about 90% of workers (Gregory and Daly, 1991). And the female-male pay ratio rose from 65% in 1969 to 85% in 1975, an extremely rapid and large increase, likely caused by the new legislation and its enforcement through tribunals (Blau and Kahn, 1995; Gregory and Daly, 1991). In the UK, equal pay legislation was implemented during the 1970–1975 period. It was initially enforced through collective bargaining agreements which removed differentiated male and female rates and, significantly, required that in workplaces with collective bargaining, women could not be paid less than the lowest male wage rate (OECD, 1988; Zabalza and Tzannatos, 1985). The interaction between collective bargaining and the equal pay law helped to sharply raise the female-male pay ratio during the 1969–1977 period by about 0.11 log points (Dolton et al., 1996). In Section 7.1 we consider possible employment effects of such dramatic alterations in relative wages.

6.3. Wage-setting institutions and the relative pay of particular groups: minimum wage laws

A final institution that directly affects the wage distribution is mandated minimum wage

coverage. Such legislation clearly impacts the bottom of the wage distribution and therefore has a disproportionate effect on low-wage workers, including youth and women. Much evidence on the impact of minimum wages looks for spikes in the wage distribution around the legal required minimum. For example, Card and Krueger (1995) find, for US teenagers in 1989 when the Federal minimum wage was $3.35/h, there was a spike at $3.35 that was the largest mass point in the teenage wage histogram; by 1991 when the minimum had risen to $4.25/h, there was again a spike in the teenage wage distribution at the new minimum wage which was higher than any other spike in the histogram. This evidence provides a prima facie case for an effect of minimum wages on the teenage wage distribution. Further, using a full-distributional simulation technique, DiNardo et al. (1996) attributed 30–70% of the 1979–1988 widening in the 50–10 log wage gap in the US to falling real minimum wages, and Blau and Kahn (1997) concluded that falling real minimum wages over this period retarded the progress of low-skill women's wages relative to low-skill men's.

Similar findings are obtained in studies analyzing the impact of minimum wages in other countries. A study by Machin and Manning (1994) of the impact of wage councils in UK for the 1979–1990 period found that, when minimum wages were raised, the distribution of pay for affected workers became more compressed. Similarly, Katz et al. (1995) concluded that rising French minimum wages in the 1980s were an important reason why France's wage distribution was stable in the face of demand shifts which were similar to those in other countries where wage inequality widened.

Other evidence that minimum wages have had an impact on the wage distribution in several OECD countries is presented in Dolado et al. (1996). For example, they show that, in France, regional wage dispersion fell dramatically when the national minimum wage was raised sharply in the 1980s. In the Netherlands, between 1981 and 1983, official youth subminimum wages were substantially lowered: for example, the minimum for 20 year olds fell from 77.5% of adult minimum to 61.5%, and for 16 year olds, the fraction fell from 47.5% to 34.5%. As a consequence, while average nominal wages rose 9% from 1980 to 1984 for those aged 23 and over who were not affected by these changes, they fell for those less than 23 years old (Dolado et al., 1996, p. 345).

In general, the impact of minimum wage mandates on wage distributions has not been the subject of much controversy. Most economists believe that a minimum wage that is binding will bring up the bottom of the wage distribution. Considerably more controversy has surrounded the issue of the employment effects of increases in the minimum wage. We consider evidence on that issue in Section 7.1.

7. Responses to labor market institutions

In the previous sections we summarized evidence suggesting that wage-setting institutions have important effects on the distribution of earnings. While wage inequality is an extremely important economic outcome, it is also important to ascertain whether there are any

employment effects of such institutions, as these may be a downside of such interventions. A recurring theme in research studying the employment effect of wage-setting institutions and employment protection measures is that other labor market responses tend to mitigate what might otherwise be even more serious employment consequences of these policies. Some of these responses were in fact designed by governments in order to avoid more negative effects of such institutions. Such policies include direct industrial subsidies, short-time compensation, fixed duration employment contracts, public employment, active labor market polices, and youth subminimum and training wages. An additional complicating factor is that there are likely to be both wage effects and direct employment effects of laws and institutions apart from those concerned with wage determination, particularly employment protection measures, industrial subsidies, unemployment insurance (UI), and active labor market policies. Finally, a serious difficulty in detecting the effects of any individual policy is that in many cases various policies move together, forming an overall package of intervention, or, in the cases of New Zealand and the UK, of deregulation. For example, in several countries, unemployment benefit levels are directly tied to the minimum wage (Dolado et al., 1996). It is important to bear this difficulty in mind as we consider specific policies below.

7.1. Employment and wage-setting institutions

Earlier, we surveyed some of the relatively fragile evidence on the impact of bargaining centralization on macroeconomic performance. While it may be difficult to say just what the macroeconomic impact of a particular bargaining regime is, larger data bases have been used more successfully to conduct more precise tests of the micro-level employment effects of wage setting institutions. Specifically, several studies have examined the impact of union pay policies, which affect relative wages, on relative employment levels. Such a research design has the advantage of differencing out common macroeconomic factors that might otherwise influence employment. Yet, if recessions or recoveries affect different groups differently, then even this strategy may not be sufficient to identify the effect of wage-setting institutions on relative employment. What is ultimately needed is sufficient time series variation to examine how particular groups fare, controlling for macroeconomic conditions.

The employment effects of several kinds of direct intervention in the wage-setting process have been examined. Some studies compare the employment outcomes for particular groups across countries and infer the effects of differences in institutional arrangements. Others examine the impact of specific policy interventions. These include changes in union wage-setting priorities as in Scandinavia, government intervention reducing union power as in the UK and New Zealand, and government directives to raise women's relative pay or to increase minimum wages in general. For these policy interventions, we observe changes in wage-setting regimes that are expected to affect relative wages and therefore possibly relative employment levels.

Overall, the evidence on the impact of wage setting institutions on employment is

considerably more mixed than the clear evidence indicating that institutions can affect the wage structure. Several papers find evidence consistent with negative employment effects of unions, but not all do. Further, the employment effects of the other types of government intervention in pay-setting appear to be either non-existent or very small in absolute value if negative, and sometimes even positive. We now discuss this evidence, beginning with the effects of collective bargaining and proceeding next to gender pay policies and minimum wages.

7.1.1. Employment effects of collective bargaining institutions

Some evidence consistent with adverse relative employment effects of compressed wage structures is provided by Blau and Kahn (1999). They compared the relative employment-to-population rates of skill groups (defined on the basis of age and education) for men during the mid-to late 1980s in several countries. In countries such as West Germany, Austria and Norway with very compressed wage structures, low skill workers had lower employment rates relative to those with middle levels of skill than in the US or the UK, suggesting negative relative employment effects of union pay compression. Similarly, Hunt (1997) finds that the reduction in the gender pay gap in eastern Germany which occurred after monetary union, likely caused by the imposition of western Germany's high union wages, resulted in relative employment declines for east German women due to layoffs.

But several studies of changes in wage distributions and employment do not support such findings. Specifically, a comparison of Canada, France and the US by Card et al. (1995) which focused on changes over time did not find evidence of a negative relationship between administered wages and employment. Their premise was that each of these countries was faced with similar shifts in demand toward more skilled labor; however, higher rates of collective bargaining coverage and more rigid wage-setting regimes in Canada and especially in France were expected to keep relative wages more stable in these countries than in the US. With downward sloping demand curves, more severe employment problems for the low skilled should have resulted in France relative to both of the other countries and in Canada relative to the US. The authors find that the relative wages of more highly skilled workers (again defined by age and education) increased over the 1980s in the US and Canada, with a slightly weaker relationship between skill level and wage gains for Canada, while the highly skilled in France stayed in the same relative position. These changes in relative wages across the three countries are consistent with expectations based on their wage setting institutions. However, relative employment-to-population ratios by skill group behaved similarly in each country, suggesting that the wage-setting institutions in France and Canada that led to more rigid relative wages than those in the US did not have the expected adverse relative employment effects during the 1980s.

Krueger and Pischke (1997) performed a similar comparison of the US and Germany over the 1979–1991 period and also concluded that wage rigidity in Germany did not appear to lead to relative disemployment for the low skilled. The low skilled did better with respect to relative wages and relative employment over the period in Germany than in

the US. Similarly, Blau and Kahn (1999) found that, among youth (aged 18–29), the less skilled in Germany had more favorable relative wage and employment levels and changes during the 1984–1991 period than those in the US, although as discussed in Section 7.3, a greater reliance on public sector employment in Germany could help to explain such a pattern.

The findings of the Blau and Kahn (1996b) on the one hand showing a negative relationship between administered wage compression and relative employment and the time-series evidence against such a pattern shown in Card et al. (1995), Blau and Kahn (1999), and Krueger and Pischke (1997) are not necessarily inconsistent. The former studied cross-sectional averages, while the latter three were primarily concerned with changes over the 1980s.[27] However, two studies of the employment effects of abrupt changes in union wage-setting policies within Sweden and Norway also find evidence of adverse employment effects, lending further support to the notion of a negative employment effect. First, Edin and Topel (1997) found strong allocative effects of Sweden's solidarity wage policy of the late 1960s and early 1970s. Specifically, relative wages in low paying industries were sharply raised during this time, and increases in out-migration from areas where these industries were located as well as decreases in relative employment in these industries were observed. Further, this relationship was strongest during the period of most intense wage compression as compared to later periods, providing some further evidence that there were negative relative employment effects of this wage policy. Recall from our discussion of Bulow and Summers (1986) in Section 3, however, that depending on how one views the allocation of labor before the wage compression, these relative employment reductions could be welfare enhancing.

Second, Kahn's (1998a) examination of the impact of Norway's wage compression during the 1987–1991 period indicated that less educated workers, whose relative wages were sharply raised, suffered relative employment declines during these years. It is possible that these relative employment changes were due to the recession which occurred at that time, rather than to the wage policy. However, during an earlier recession period in which bargaining had become less centralized (1980–1983), the relative wages of less educated workers declined and their relative employment levels actually increased among men, while remaining constant among women. Again, the "differences in differ-

[27] Blau and Kahn (1999) did find that relative wage and employment levels were both higher for German than for US less educated youth in 1984 and in 1991, while Blau and Kahn (1996b) showed that low skilled employment to population ratios were relatively lower among men in general in Germany in the 1980s than in the US. Part of the explanation for this discrepancy may lie in the fact that low skill German youth were much more likely to work for the public sector than their American counterparts. Nickell (1997) shows that relative spending on active labor market policies in Germany during the 1989–1994 was about 8 times as high in Germany as in the US (this was defined as spending per unemployed person as a percentage of GDP per member of the labor force). To the extent that such policies are disproportionately directed at youth and provide public sector jobs, they may help to account for the relative success of German youth. Finally, as pointed out by Nickell and Bell (1996a), less educated German youth have greater cognitive skills than those in the US, a difference that could also explain the outcomes for wages and employment.

ences" framework provides some support for the notion that there are negative employment effects of union wage policies.[28]

Two cases in which the government passed laws which were apparently designed to reduce union power, the UK in the 1980s and New Zealand in 1991, enable us to examine the impact on employment of reductions in the extent of collective bargaining. These results, however, are considerably less clear-cut than those for Sweden and Norway in which union policies increased wage compression.

In the UK, the Thatcher reforms constituted a many-faceted program designed to move the economy toward a laissez-faire ideal (Blanchflower and Freeman, 1993), including a variety of policy interventions such as abolishing closed shops and limiting union picketing, as well as reducing the generosity of the welfare state by lowering the UI replacement ratio, and abolishing wages councils. The Thatcher programs appear to have had a strong negative effect on union coverage in the UK (Freeman and Pelletier, 1990). And the responsiveness of wages and employment at the micro-level to demand changes did increase as a result of this package. However, these developments were not successful in lowering unemployment generally or in raising transitions out of unemployment (Blanchflower and Freeman, 1993). Union relative wage effects remained at about 10%, a relatively high level by international standards (see above), implying that insider power was still a force to contend with (Blanchflower, 1996).

In contrast to the UK experience, the New Zealand Employment Contracts Act of 1991, which also substantially reduced union power, appears to have had some positive employment effects. Specifically, this legislation outlawed compulsory unionism and abolished national wage awards. Since the new legislation was implemented at different rates across different industries due to different contract expiration dates, one can estimate its effect by using the industries that had not yet implemented the changes as a control group. Using this research design, Maloney (1994) found that the law sharply reduced union coverage and raised employment. However, relative wages were unaffected, implying that fringe benefits or work rules changes were the mechanism for increased employment.

The findings for the UK and New Zealand leave us with some puzzles. In both countries, government reforms appear to have reduced union density without affecting union relative wages. Yet in New Zealand, there were positive employment effects of such policies, while in the UK there did not appear to be any employment effects. Progress in resolving this contradiction might be made if we were able to learn more about firms' responses to the new legal environment, including impacts on fringe benefits and work rules, as well as the entry of new non-union businesses.

7.1.2. Employment effects of gender pay policies

As we gave seen, some types of government policy to address gender discrimination take the form of direct government intervention in the wage-setting process which may have

[28] The fact that the 1980–1983 recession was weaker than the 1987–1991 slump could have contributed to the worse employment outcome for less educated workers in the latter period.

effects on relative employment. Of particular relevance is comparable worth where the government mandates that employers compensate workers equally for work of equal value to the enterprise regardless of occupation. This policy is currently in effect nation-wide in Australia, in Ontario Province (Canada) for virtually all employers and in about 20 states in the US for certain public sector employees. It recognizes that men and women are to a considerable degree segregated in different jobs and that a simple policy of equal pay for equal work will leave a substantial portion of the gender gap in pay untouched.[29] As we saw in Section 6.2, this policy appears to have been quite successful in raising the relative pay of women in a fairly short period of time through a realignment of pay in predominantly male and female jobs. However, by increasing wages of occupations in which women are disproportionately employed, this policy runs the risk of lowering women's relative employment levels, at least assuming downward sloping labor demand curves.[30]

Economists have studied the employment effects of these comparable worth interventions in the cases of Australia and for government employees in San Jose (California), Minnesota and Washington State. In each case, comparable worth had a noticeably positive effect on women's relative wages, particularly in Australia (Gregory and Duncan, 1981; O'Neill et al., 1989; Killingsworth, 1990). And in several cases, the implementation of comparable worth came at a time of rising relative female employment, which continued to rise even after women's relative wages were increased. However, when each of these analyses controlled for the overall growth of women's employment, it was found that women's relative employment grew more slowly than it otherwise would have, although the effects are not large. One interpretation of these results is that the negative employment effects were not sizable enough to outweigh growing female labor market attachment or, in the cases of public employees in the US, to cause layoffs of female government workers.

In contrast to the research on comparable worth, which finds some evidence, albeit modest, of employment losses in response to administered wages, Manning's (1996) study of the impact of the UK's Equal Pay Act of 1970 and Sex Discrimination Act of 1975 shows no employment losses for women, despite their large gains in relative wages. Manning (1996) finds that during this time, wage changes and employment levels by industry were strongly positively related for women, but less so for men, a pattern he interprets as being consistent with the notion that women face more monopsony power than men do. And the positive relationship became slightly less so after 1971, implying that monopsony power fell after the passage of the Equal Pay Act of 1970. Manning's (1996) results provide an explanation for why female relative employment did not decline in the face of large exogenous increases in their relative pay. As we shall see below, some results for the impact of minimum wage laws in the US follow a similar pattern, again suggesting monopsony elements in the labor market.

[29] For further discussion of the issues in implementing comparable worth, see Gregory and Daly (1991), Killingsworth (1990) or Gunderson and Robb (1991).

[30] Below we discuss an alternative view of employment responses to mandated wage increases in the presence of employer monopsony (Manning, 1996).

Differential monopsony power facing men and women could explain the existence of gender discrimination (Madden, 1973), as well as Manning's (1996) results of different wage-employment relationships for men and women. Although women's labor supply is in general more elastic to the economy than men's (implying that men might face more monopsonistic exploitation), Manning (1996) argues that women's supply of labor to individual employers may be less elastic because women are for family-related reasons likely to be less mobile. While this appears plausible, this factor may or may not outweigh women's greater overall wage elasticity of labor supply.

Findings from studies of quitting in the United States do not support the idea that men's labor supply at the firm level is more sensitive to wages than women's. For example, calculations based on results presented in Viscusi (1980) and Blau and Kahn (1981) on quit behavior of men and women in the 1970s indicate that the derivative of the probability of quitting with respect to the log of pay is more negative for women than men, in contrast to the monopsony story. And results in both papers suggest that the elasticity of quitting with respect to the wage was similar for men and women. Further, Light and Ureta (1992) found for a later period that even the wage elasticity (as well as the derivative) of the quitting hazard was more negative for women than for men.[31] These studies indicate that we must look elsewhere for explanations of gender discrimination as well for the differences in the wage-employment relationship by gender that Manning (1996) found. Of course, Manning's basic finding that women's employment did not suffer despite the massive increase in their relative wages remains consistent with the idea that women face monopsony power.

7.1.3. Employment effects of minimum wages

Minimum wage interventions are a final type of wage-setting device that has been much studied in the US and increasingly in other countries. Generally, the minimum wage is too low to affect major portions of the labor market, although specific subgroups such as teenagers may be more directly affected than workers in general by legislated minimum wage increases. Most of the research on the impact of minimum wages finds little evidence of negative employment effects, and when these have been found, they are generally too small to have an important effect on the labor market.[32]

[31] Meitzen (1986) also analyzed US data on male and female quitting. However, he included several measures of pay (e.g., starting wages, top pay level in the job, market wages) and men's and women's relative responsiveness differed by type of pay variable; thus his results do not allow one to determine which group's labor supply is more sensitive to wages overall.

[32] An exception is Castillo-Freeman and Freeman's (1992) study of the US decision to bring minimum wage coverage to Puerto Rico starting in 1974. By 1988, about 28% of workers in Puerto Rico were paid within 5 cents of the US minimum wage of $3.35/h; on the mainland, roughly 25% of teenagers were paid within 5 cents of the minimum at this time (Card and Krueger, 1995). And in 1987, the US minimum wage was about 63% of the average manufacturing wage in Puerto Rico but only 34% on the mainland. Thus, the high minimum in Puerto Rico had the potential to greatly disrupt its labor market. Castillo-Freeman and Freeman (1992) in fact find large disemployment effects, but Krueger (1995) finds that this result is very sensitive to econometric issues such as weighting.

Early research in a time-series framework for the US, based on minimum wage changes in the 1960s and 1970s, found that, for teenagers, a 10% increase in the minimum wage led to a 1–3% fall in teenage employment. However, this research has been criticized on the grounds that the measure of minimum wage changes was itself negatively confounded with overall demand changes, since it included average wages in the denominator. This could have induced a negative estimated employment effect even if in fact minimum wages had no impact on employment (Card et al., 1994; Card and Krueger, 1995). More recent research looks for appropriate control groups against which to compare the employment changes of teenagers or other low wage workers who are most likely to be affected by minimum wage increases. Much of this work finds either no effect or in some cases a positive effect on employment of increasing the minimum wage (Card and Krueger, 1995). Some recent research on the US minimum wage continues to find negative effects, including Neumark and Wascher (1992) and Deere et al. (1995), although only the latter study reports a large disemployment effect. Specifically, Deere et al. (1995) find that after the 1990 and 1991 minimum wage increases in the US, youth employment fell sharply relative to adults. However, such an outcome could also have been explained by the recession of the early 1990s having a disproportionate negative effect on youth. The finding of a positive effect on employment of increasing the minimum wage has been interpreted as possible evidence of employer monopsony, as have Manning's (1996) findings for the UK sex discrimination policies described above. Of course, a zero effect on employment is also consistent with employer monopsony, but not with the ordinary competitive demand model.

Minimum wage research on other countries has found evidence of negative effects in some cases, with little reported impact in other instances. For example, Abowd et al. (1999) found that French minimum wage increases in the 1980s lowered the employment of workers at the minimum relative to a control group just above the minimum. The effects on workers at the minimum were large, but since they comprised a small portion of the labor force, the total effects on employment were small. However, an alternative analysis of France during the 1967–1985 period in which minimum wages were substantially raised finds that employment growth was actually higher in regions most affected by minimum wage increases (Dolado et al., 1996). While such a finding could have been caused by the relocation of businesses to low wage regions (a long run adjustment that could occur even if mininum wage increases affected these regions disproportionately), it does not suggest a negative employment impact of raising the minimum wage.

Some evidence compatible with negative employment effects of the minimum wage comes from the Netherlands where, from 1981 to 1983, youth submininum wages were sharply lowered. Dolado et al. (1996) do find that relative to youth employment changes in the economy overall, youth relative employment generally rose in low paying occupations, the ones most likely to be affected by minimum wage changes. However, as the authors note, with the aggregate-level data used, it is not clear whether such differences were statistically significant. In addition, if changes in the Netherlands' minimum wages had an effect on youth employment, it must be the case that employers were taking advantage of

the youth subminimum. Yet previous work on the US and the UK finds that employers do not seem to use the subminima in those countries (Katz and Krueger, 1992; Machin and Manning, 1994). If employers in the Netherlands also in general did not utilize the youth subminimum, then the findings reported in Dolado et al. (1996) must have been caused by some factor(s) other than the change in the minimum wage law.

Evidence consistent with a negative effect of the minimum wage was also found for Spain which sharply raised the minimum wage for youth 16 years and under in 1990. Dolado et al. (1996) find that this policy led to a substitution of adults for youths but, paradoxically, also to increasing total employment. Mixed evidence is available for Canada based on the traditional time series approach. Summarizing Canadian research, Card and Krueger (1995) report that while negative effects on teenage employment were obtained for the 1956–1975 period, the effects were statistically insignificant for 1976–1988, with a negative point estimate for males and a positive one for females. Finally, Machin and Manning's (1994) findings on the impact of wages councils on employment in the UK are consistent with the recent US findings of zero or even positive employment effects of minimum wage increases.

7.1.4. Summary

The evidence on the impact of wage-setting institutions on employment is mixed. On the one hand, there is some evidence which is consistent with adverse relative employment effects of compressed wage structures due to union wage setting, although not all studies obtain this result. In contrast, comparable worth and minimum wages do not appear to have dramatic disemployment effects. On its face, this collection of evidence is more consistent with the monopsony model under which employment may increase, decrease or remain unchanged in the face of a wage increase, than with the traditional competitive model. Of course it may be easier to uphold a theory in which any outcome is possible (monopsony) than a theory that has only one prediction for the direction of employment effects. Recent approaches to such models which emphasize that an upward sloping labor supply schedule may characterize any firm that must expend resources in order to recruit labor are likely to be more palatable to many economists than the old view which required the implausible assumption of a labor market controlled by one employer. Nonetheless, at present the jury is still out as to whether in detail this set of results fits the monopsony model, i.e., do the cases in which higher relative wages are associated with decreases or increases in relative employment form a coherent pattern? How do we reconcile the broad findings of employment effects for union wage compression, particularly the before and after evidence for Sweden and Norway, with the instances where unions were not found to lower employment and with mixed findings for comparable worth and minimum wages? Even taking the latter two separately, there are some puzzles. Many minimum wage increases are relatively small, plausibly falling in the region where, under monopsony, relative employment would rise, or at least fail to decrease. However, comparable worth and other interventions described above in instances raised gender wage ratios substantially over a short period of time. Might we in these cases expect firmer evidence of

negative relative employment effects even under monopsony? Finally, it cannot at this point be ruled out that statistical problems, such as endogeneity of institutional intervention or overly short time series, will in the end explain the failure to uncover negative employment responses in some cases.

7.2. Labor market flexibility and employment responses

In exploring possible reasons for persistently high European unemployment, many economists have focused on the issue of labor market flexibility, the openness of the labor market to adjustments in the face of shifting supply and demand. It has been hypothesized that, in Europe, rigid institutions governing wage-setting and the allocation of labor prevented adjustment to the demand shocks of the 1970s and early 1980s (Siebert, 1997). And, forces of hysteresis – in which initial increases in unemployment become self-reinforcing – served to keep unemployment high even after the initial shock had passed.[33] An example of a mechanism which has been proposed as causing hysteresis is a deterioration of workers' skills as the duration of unemployment increases; this lowers their probability of leaving unemployment. Or, as another example, it is claimed that, at a given unemployment rate, unions will be more aggressive the higher unemployment was in the previous period, since insiders' jobs are less threatened with falling than with rising unemployment. Thus, in principle, union wage bargaining slows down the fall in unemployment during a recovery (Layard et al., 1991).

The issue of flexibility took center stage in the OECD's (1994b) analysis of the European employment problem. Among other recommendations, the OECD advocated a program of increased flexibility in work time, making wages more responsive to local labor market conditions, reducing the intrusiveness of employment security provisions, and reforming UI systems to reduce long-term unemployment. Implicit in these recommendations was the assumption that lack of flexibility was inhibiting the growth of employment in Europe. We now consider evidence on the employment impact of differences and changes in the various dimensions of labor market flexibility, in an attempt to determine whether the OECD's concerns were well-founded. Particular aspects of European labor markets that have been hypothesized to create higher unemployment include employment security measures, UI systems, payroll taxes, and industrial subsidies.

One of the most contentious issues concerning high European unemployment is the degree of mandated job protection, usually given in the form of required advance notice of layoffs, generous severance pay, or required negotiations with unions or works councils over downsizing. As discussed in Section 4, the effect of such restrictions on average employment is theoretically ambiguous, although they are expected to slow the adjustment of employment to both positive and negative demand shocks. The limited empirical evidence on this question does suggest that these restrictions may have some negative effects on average employment, although not all analyses obtain this result.

[33] See, for example, Blank and Freeman (1994), Blanchard and Summers (1986), or Layard et al. (1991).

First, Lazear (1990) used a cross-section time-series approach to study this issue, analyzing data on 22 countries between 1956 and 1984. In a model without country effects, he finds that more generous mandated severance pay lowers the employment-to-population ratio, the labor force participation rate, and weekly hours per worker, with ambiguous effects on unemployment rates. Much of this pattern is due to between country effects, but when country dummies are added to the model, the effect of severance pay on the employment-to-population ratios remains negative, although it is no longer significantly different from zero; the effect on hours per worker remains significantly negative. The results when country dummies are included are especially important because they cannot be due to permanent country effects. Lazear's findings appear to indicate that more generous mandated severance pay reduces new hiring and transforms what would have been full-time jobs into part-time jobs, presumably because the mandate does not apply to part-time workers.

Second, in a reanalysis of Lazear's (1990) data with what the authors claim are corrections of some erroneous data series, Addison and Grosso (1996) confirm the negative effects of severance pay on the employment-to-population ratio and the labor force participation rate; however, they find that more generous severance pay raises weekly hours. These findings are not affected by the inclusion of country dummy variables. While Addison and Grosso (1996) find a significantly positive effect on unemployment as well, it is not robust to the inclusion of country dummies. The positive hours effects may indicate that employers treat mandated severance pay as a fixed cost of employment, although the comparison with Lazear (1990) indicates that estimates of the effect of severance pay on hours per worker are somewhat fragile. Finally, Addison and Grosso (1996) find that stricter notice requirements raise employment, labor force participation and hours, and lower unemployment, even in fixed effects models. In contrast, Lazear (1990) only reported results for required notice without controlling for country dummies and found adverse effects on employment, hours, and labor force participation.

A cross-country study by Nickell (1997) also finds little association between employment protection and overall unemployment, but suggests that the long-run effects of such policies may be more adverse than the short-run effects. He examined the impact of employment protection on average unemployment over the 1983–1988 and 1989–1994 periods. His measure of employment protection was a ranking of the degree of intrusiveness of government policy, with the US having the least protection and Italy the most among the 20 countries studied. While Nickell was not able to control for country fixed effects, he did include controls for UI system characteristics, union bargaining coverage and coordination, extent of active labor market policies, and the change in inflation. He found that employment protection had small and statistically insignificant effects on overall unemployment. Interestingly, however, he obtained large positive and nearly significant effects on long term unemployment and large nearly significant negative effects on short term unemployment. This contrast between the effect of employment protection on long-term and short-term unemployment is revealing, since the long term effect probably reflects reduced new hiring, while the short term impact is probably due to the increased

protection afforded insiders. These two impacts roughly canceled, leading to no association between the overall unemployment rate and employment protection. An interesting extension of this approach would be to examine the effects of notice and severance pay separately on long term and short term unemployment; such an exercise could shed light on Addison and Grosso's (1996) contradictory findings for the impact of notice and severance pay on overall employment.[34]

Further evidence on the effects of mandated employment protection comes from variation across states in the US in the degree of employment protection. This source of variability has been found very useful in studies of other labor market policies, including minimum wages, unemployment insurance, and laws regulating collective bargaining since some states can serve as controls for others. And, as Ehrenberg (1994) points out, this may be a particularly fruitful research strategy for detecting the impact of labor market interventions since there may be fewer unmeasured differences across states in the US than across certain countries in the OECD.

Between 1980 and 1989, the number of states in the US in which employees had the right to sue their employer for wrongful discharge increased from 13 to 45. Dertouzos and Karoly (1993) exploit this variation to estimate the impact of the legal environment on employment. Using a state fixed effects approach and instrumenting for the enactment of such job protection, the authors find that limiting employers' right to unilaterally fire workers reduces state employment levels by a statistically significant 1–3%, all else equal. This may not indicate a fall of 1–3% in work hours. Since such protection constitutes a fixed costs of employment, we expect hours per worker to increase, perhaps offsetting some of the employment declines. Unfortunately, the authors do not examine this issue. The Dertouzos and Karoly study is especially noteworthy because it is the only one that even attempts to take into account the endogeneity of the enactment of employment protections. There may be a negative bias built into single equation estimates of the impact of such legislation on employment in that the public is likely to be most interested in enacting job protection when many people are losing their jobs. On the other hand, it is possible that the extension to employees of the right to sue for wrongful discharge occurred at the same time states enacted other regulations that lowered employment, thus inflating the coefficient on employment protection in Dertouzos and Karoly's estimates.

In summary, research examining the impact of employment protection has found some support for the idea that more generous mandated severance pay (or wrongful termination penalties in the case of the US) lowers overall employment. But the effects on unemployment rates are considerably more mixed, and high European unemployment is the major

[34] Nickell (1997) also found a negative, significant association between employment protection and the overall employment to population ratio. However, he argued that this was due to the fact that employment protection was strong in countries such as Spain and Italy, where women's labor force participation rates are relatively low. Consistent with this reasoning and with his overall unemployment results, he found that the male employment to population ratio for those age 25–54 was not significantly related to the strength of employment protection policies.

labor market feature that groups like the OECD have been attempting to explain. Further, it is possible that any positive effects of the generosity of mandated severance pay on unemployment rates estimated in cross-sectional data are biased upward by reverse causation: higher unemployment could lead to demands for more generous severance pay; as the position of insiders becomes more threatened, they have a greater interest in policies to protect themselves (Saint-Paul, 1996). Nickell's (1997) finding of a positive effect of employment protection on long-term unemployment that roughly cancels out its negative effect on short term unemployment can perhaps, in an accounting sense, explain the finding of a lack of effect on the overall unemployment rate. Further, his findings that protection reduces overall employment but not male employment suggests a difference between the estimates of the effects of employment protection on employment and on unemployment. This contrast deserves further study, with labor force participation being the center of attention.

Taken together this evidence suggests that employment protection is indeed a possible contender for explaining low European employment-to-population ratios, although results for overall unemployment are not found to be strong. The provisions of unemployment insurance systems are another area where there are significant differences between Europe and the US. Specifically, European UI systems traditionally had higher replacement rates and much longer potential durations than those of the US. There is abundant evidence from US studies that more generous UI benefits lengthen unemployment duration (see, e.g., Ehrenberg and Smith, 1997); and long-term unemployment is much more prevalent in Europe than in the US (Burtless, 1987; Nickell, 1997).[35] Additional evidence suggesting that Europe's UI systems may contribute to their unemployment problems comes from cross-country studies. Addison and Grosso (1996) find that, other things equal, longer mandated UI benefit duration significantly raises a country's unemployment rate, while having significantly negative effects on employment and labor force participation rates. And Nickell (1997) also finds some cross-sectional evidence that UI system generosity contributes to unemployment. The UI replacement rate had positive effects on total, short-term and long-term unemployment; and the effects for total and short-term unemployment were statistically significant. The maximum UI benefit duration also had positive effects, but only one, the impact on long-term unemployment, was statistically significant.

These results therefore imply at least some role for UI benefits in explaining Europe's relatively high unemployment rates. However, some qualifications are in order. As was the case for mandated protection, the generosity of UI systems can also respond to actual unemployment, implying reverse causality. For example, in the early 1990s, the US mandated increased UI benefit duration in response to the relatively high unemployment rates at the time (Social Security Administration, 1991). Moreover, before 1973, the US had far less generous UI than other OECD countries, as well as less prevalent collective bargaining, and less intrusive employment protection, and yet it had a relatively high unemployment rate. Thus, both the more flexible labor market institutions in the US

[35] However, Burtless (1987) points out that parameter estimates from such research are too small to fully explain the longer European unemployment duration in comparison with that in the US.

and the more rigid institutions of the other OECD countries have at times been compatible with lower unemployment rates. If the more regulated labor markets of the European countries are responsible for their currently more severe employment problems, it is likely that this is because, as we suggested at the outset of this section, they prevented adjustment to the significant demand shocks of the 1970s and early 1980s, with forces of hysteresis working to keep unemployment high even after the initial shocks had subsided.

Although high European taxes have been cited as reducing labor demand there (OECD, 1990), it appears that, unlike job protection and UI systems, payroll taxes and mandated benefits do not provide a significant explanation for Europe's employment problems. Payroll taxes and mandated benefits can have adverse employment effects if it is not possible to fully shift them back to labor in the form of lower wages. Europe's high wage floors make this a distinct possibility. However, while the OECD made this argument for the short run, it found complete shifting of payroll taxes onto labor in the long run (OECD, 1990). Moreover, although Nickell (1997) found that the impact of total labor taxes (payroll, income and consumption) on unemployment was positive and statistically significant, the estimated effect of a politically realistic change in taxes was quantitatively small. Nickell's measure of payroll taxes included mandated benefits, as the tax rate is the ratio of total labor costs to wages, and most European fringe benefits are mandated.

In the face of at least some evidence that labor market regulation in the form of job protection and UI benefits have contributed to low European employment, it is noteworthy that many European countries have attempted to make their labor markets more flexible. Yet these efforts have not succeeded in lowering European unemployment. Specific measures instituted in France, Italy, Spain, and Sweden over the 1980s and 1990s include expansion of fixed duration contracts, reduction of severance pay, reduction of sick leave benefits, and allowing greater use of youth subminimum wages; in addition, as we have seen, the UK greatly restricted union power and made UI benefits less generous relative to wages during this time. Yet in each of these countries except the UK, the government has regressed along some dimensions and introduced further regulations into the employment relationship. In some cases, firing has been made more difficult (Sweden in 1985, Italy in 1990 or France in 1989), while in others, fixed term contracts have been made more difficult to use (Spain in 1994).[36]

Bertola and Ichino (1995) argue that these instances of government backsliding reduce the credibility of the flexibility reforms. When the government does permit easier firing, for example, this has the immediate effect of allowing layoffs; however, for the countries other than the UK, firms and unions anticipate further protections to be enacted, making wage demands by current insiders more aggressive.[37] And firms refrain from hiring now because of anticipated higher future firing costs. Thus, with more flexibility that is not

[36] We will return to issue of fixed duration employment contracts as a response to the rigidities of European labor markets.

[37] Friesen's (1996) study of the wage effects of mandated severance pay and advance notice requirements in Canada found that insiders covered by such protection, particularly union members, were able to extract higher wages than otherwise. This finding supports Bertola and Ichino's (1995) claim.

believed to be permanent, we get the worst of all worlds – more layoffs but not much change in new hiring. In contrast, the authors argue that the UK's reforms were unambiguously and permanently in the direction of more flexibility (the UK "crossed the river"); and in that case, flows both into and out of unemployment became generally higher than those of other European countries, although not nearly as high in the US. Thus, Thatcher's programs did lead to more flexibility in the UK and its long term unemployment rate fell between 1983–1988 and 1989–1994 (Nickell, 1997). However, by the mid-1990s, the UK continued to have "European" rather than "American" unemployment rates, indicating that the Thatcher programs had not yet brought the UK to low unemployment (Blanchflower and Freeman, 1994).

While increased flexibility is one type of policy reform on the table in Europe, another one is employment subsidies. Nickell and Bell (1996b) express some doubts about the potential effectiveness of subsidies, arguing that they will be passed onto wages like payroll taxes are. Leonard and Van Audenrode (1993) argue that not only are industrial subsidies much more extensive in Europe than in the US; they are also given disproportionately to declining firms and industries. Thus, declining firms are kept going by taxing expanding firms. An important indirect effect of such policies, according to Leonard and Van Audenrode (1993) is to make European unions more aggressive in bargaining, since they know that any extreme disemployment effects of their wage bargains will be reduced through subsidies. The authors use firm-level data on Belgium that show these predicted effects in fact occur, even after instrumenting for the receipt of subsidies by the firms in question. Thus, a case can be made that subsidies may lower employment through excessive union wage demands; however, it is not possible to know the degree to which such effects, if any, can account for European-US unemployment differences. All we can say is that subsidies are more prevalent in Europe, as are unions, employment protection, and social benefits generally.

A further implication of the prevalence of subsidies concerns the impact of executives on the value of the firm. In settings where failing firms are likely to receive income subsidies paid for by taxes on successful firms, the value of the firm's assets is less dependent on the actions of its chief executive than if there were no subsidies. This reasoning could explain Abowd and Bognanno's (1995) finding that US executives earn higher total compensation levels than those in Europe. That is, in the US, stockholders gain more when executives do a good job and lose more when they make mistakes than in countries with more subsidies.

An important final point to note regarding the impact of these types of labor market institutions on employment is that such policies may plausibly affect the structure of employment as well as its level. This was implicit in our earlier discussion of the employment effects of wage-setting institutions where we found some evidence of adverse relative employment effects on less skilled workers of institutions which compress wages. In a recent paper, Davis and Henrekson (1997) raise a related issue of the impact of the policy environment on the firm size and industry distribution of employment. Their study focuses on Sweden which they characterize as having a set of economic policies, including busi-

ness taxes, credit market regulations, as well as the national pension system, employment security laws and other wage setting institutions, which strongly disfavor less capital-intensive and smaller firms, as well as entry by new firms and individual and family ownership of business. Consistent with this, they find that, in comparison to the US, Sweden's employment distribution in the mid-1980s was sharply tilted away from low-wage industries[38] and industries with greater employment shares for smaller firms. They also found that Sweden had an unusually high share of employment in large firms compared to other European countries and its self-employment rate was the lowest of all the OECD countries. These findings suggest that labor market institutions may affect the structure of demand for labor along a number of dimensions.

7.3. Countervailing institutional responses to adverse labor market consequences of institutions

While we have uncovered some evidence that particular European institutions may be partly responsible for high European unemployment, one might have feared far worse impacts of high wage floors, rigid wage structures, high firing costs, extremely generous UI benefits, and the like, on labor market performance. In fact, there are many responses by governments in these countries that serve to limit the adverse effects of the system of social protection we have described. These include short-time compensation, public employment, active labor market policies, youth subminimum and training wages, and fixed duration employment contracts. In addition, in some countries, particularly Italy and Spain, there is a very large "unregulated" sector that can serve as an outlet for workers and firms shut out by the high cost of doing business by the rules. This includes both those employed in the "underground" economy where employment regulations are flouted and the self-employed who are by and large exempt from these regulations. European labor markets may thus be more flexible in practice than they would appear at first blush.

One mark of flexibility is the degree to which labor inputs are allowed to vary in response to changes in demand. And job creation and destruction are both much less rapid in Europe than in the US, implying less flexible labor allocation in Europe. However, Abraham and Houseman (1994) and Van Audenrode (1994) note that in several European countries, workers can much more easily collect short-time compensation from the government than is the case in the United States. An implication of this difference is that in Europe, hours per worker adjustments are in fact more cyclically sensitive than in the US, in contrast to the greater sensitivity of employment to demand in the US. The result of these offsetting patterns is that, in the 1980s, the adjustment of total production worker hours was similar in West Germany, France and Belgium to that in the US (Abraham and Houseman, 1994).

The European practice of hours flexibility may actually provide more income insurance

[38] This is consistent with Edin and Topel's (1997) result discussed above that found decreases in relative employment in low-wage industries in response to Sweden's solidarity wage policy of the late 1960s and early 1970s.

than does the US practice of employment adjustments, since, in the former, a 10% cut in total labor input, for example, gets shared across all workers, while in the US, it is more likely to be concentrated on those who become laid off.[39] And, of course, as discussed above, European UI benefits are more generous than those in the US. Further, the greater incidence of national health insurance in Europe implies that finding work is less important for obtaining health care there.[40] Therefore, while unemployment is much more prevalent in Europe than in the US, it appears to have less severe consequences for poverty.

While Europe appears better able to tolerate high unemployment, it is also true that European governments intervene to a greater extent in order to shore up employment than is the case in America. The public sector is larger in several European countries than it is in the US, and many of these countries spend considerably larger amounts per unemployed worker (in relation to output per worker) on training and relocation programs than is true in the US.[41] While the size of the public sector undoubtedly reflects the electorate's demand for publicly-produced services, in some countries, the government has explicitly been used to provide employment for those out of work. For example, in Sweden and Norway, government employment of unskilled workers has been found to increase during periods of wage compression in which the wages of workers at the bottom of the distribution have been raised the most. This government hiring may serve to limit the disemployment effects of union wage bargaining (Björklund and Freeman, 1997; Edin and Topel, 1997; Kahn, 1998a). Similarly, in the late 1980s, government employment has been found to be more prevalent, both absolutely and relative to other groups, among less-skilled youth in Germany than among such young people in the US (Blau and Kahn, 1999). Again, in Europe, the group most likely to be shut out by high union wages for the low end of the wage distribution was more likely to find government employment there than in the US.

Active labor market policies are another corrective mechanism to unemployment problems, although detecting an effect of such policies is difficult, since they arise when unemployment becomes a problem. Nickell (1997) found across 20 OECD countries in the 1980s and 1990s that spending on active labor market policies had negative and significant effects on overall and long-term unemployment that were of moderate size.[42] Thus, these programs may raise the likelihood that unemployed workers will eventually find work. However, these programs were also found to have insignificant (though positive) effects on the overall employment to population ratio and negative, insignificant effects on total hours worked. Thus, the evidence for a favorable effect of active labor market programs is not strong overall. The ambiguity of the findings may by due to the fact

[39] This point applies most strongly if layoffs are randomly distributed. In fact, in the US, layoffs are distributed by inverse seniority. So the European system could actually increase the probability of income loss for more senior workers, while providing considerably more income insurance for junior workers.

[40] This is in fact likely to be an additional reason (besides less generous UI systems) for a shorter duration of unemployment in the US.

[41] For discussions of public sector employment, see Björklund and Freeman (1997), Edin and Topel (1997), Blank (1994), Kahn (1998a), Blau and Kahn (1999). Nickell (1997) shows that, for the 1989–1994 period, relative spending on active labor market programs in the US was last out of 20 OECD countries.

[42] Active labor market policies were instrumented, in part due to the endogeneity argument noted above.

that the effectiveness of such programs is limited by two further responses that they appear to foster. First, there is some evidence that they crowd out private sector employment, so that some people taking part in such programs might have found other work in their absence (Forslund and Krueger, 1997). Second, it is possible that unions will become more aggressive in wage bargaining if they know that unemployed union members can get training, relocation allowances or public employment (Calmfors and Forslund, 1991).

While much of the European employment relationship is highly regulated, new forms of flexibility have emerged that provide more job opportunities than otherwise. Specifically, in countries like Italy and Spain, there is a large unregulated sector in which taxes, union wages and other rules can be avoided. In some cases this takes the form of self-employment as in Italy, while in others, the government does not enforce its regulations (Erickson and Ichino, 1995; de la Rica and Lemieux, 1994). Perhaps at least as important as the underground economy is the sharply increasing practice in the OECD of allowing fixed duration employment contracts. For example, the fraction of European Community workers with temporary jobs grew from 4% in 1983 to 10% in 1991 (Bentolila and Dolado, 1994).

While it is tempting to view the growth of temporary employment in the OECD as at least in part a response to the rigidity of the regular employment system, it is important to bear in mind that there are a number of demand and supply-side reasons for the use of such arrangements, and, while data are limited, indications are that the incidence of such employment has been increasing in the US as well (Houseman, 1997). A comparison with the US would be instructive in indicating whether the other OECD countries have been especially motivated to turn to these arrangements by their less flexible labor market institutions. Unfortunately, there are two difficulties which limit the usefulness of such a comparison. First, it is not possible to obtain US data for precisely the same time period or for the same definition of temporary work as for the other OECD countries. Second, the incidence of such jobs in some OECD countries has been restricted by regulations which ban certain types altogether or heavily regulate the conditions under which other types may be offered (OECD, 1993).

Bearing these qualifications in mind, in Table 4, we show data for 1983 and 1991 for the other OECD countries based on a common definition of temporary work, including those employed by temporary agencies and with fixed-term contracts, and data for 1995 for the US based on two alternatives which match up as closely as possible with the OECD definition. The first includes agency temporaries and short-term hires, i.e., workers who said their jobs were temporary. The second definition additionally includes on-call workers.[43] Table 4 indicates that the incidence of temporary employment in the US, 4.5–6.4%,

[43] The definitions are mutually exclusive, so short-term workers do not include those who were employed by temporary help agencies or were on call. They also do not include those who were independent contractors or worked for a company that contracted out services. The latter two categories comprised 1.0 and 1.3% of workers, respectively (Houseman, 1997). Note that the non-US averages shown in the table differ from those cited in the text above from Bentolila and Dolado (1994) for the EC because the latter are employment weighted and include only the EC countries.

Table 4
Temporary employment as a percent of total employment in OECD countries, 1983 and 1991[a]

	1983	1991
Belgium	5.4	5.1
Denmark	12.5[1]	11.9
Finland	11.1	13.1
France	3.3	10.1
Germany (W)	9.9[1]	9.3
Greece	16.3	14.7
Ireland	6.2	8.3
Italy	6.6	5.4
Luxembourg	3.2	3.3
Netherlands	5.8	7.7
Portugal	16.9[2]	16.5
Spain	11.3[3]	32.2
UK	5.5	5.3
Australia	21.1[2]	19.7
Japan	10.3	10.5
Non-US average (unweighted)	9.7	11.5
US		
Agency temporaries and short-term hires	na	4.5[4]
Agency temporaries; short-term hires; and on-call workers	na	6.4[4]

[a] Source: For the US: Houseman (1997) based on CPS data; for other countries: Bentolila and Dolado (1994), based on OECD data. Notes: 1, 1984; 2, 1987; 3, 1985; 4, 1995. With the exception of the US, temporary employment includes those employed by temporary agencies and on direct fixed-term contracts.

tends to be lower than in many of the other OECD countries, and less than the unweighted average for the others of 11.5% in 1991. This is especially impressive given that the US data are for 1995 and temporary employment has been increasing. Moreover, as mentioned earlier, many of the other countries heavily regulate the use of temporary employment which is not the case in the US. Thus, the data in Table 4 provide some preliminary support for the notion that the more rigid labor market institutions in the other OECD countries have provided additional impetus for the growth of temporary employment there.

Such arrangements can, however, have mixed effects on employment in the European context. On the one hand, temporary workers can obviously be more easily terminated than permanent workers, and their wages generally are lower as well, providing some

additional job opportunities. On the other hand, temporary employment may also contribute to the dualization of many European labor markets, as insiders become more insulated from economic fluctuations by temporary workers. For example, in a study of Spain's experience with legalizing temporary employment contracts as of 1984, Bentolila and Dolado (1994) found that in fact employment did become more cyclically sensitive as a result, and long-term unemployment fell. However, real wage growth of bargained wages actually accelerated after 1984 (despite extremely high unemployment rates of over 20%), a development the authors interpret as evidence that insiders have become more aggressive. This conclusion is bolstered with firm level data that show the same effect: temporary workers appear to act as a buffer between permanent workers and unemployment. Thus, while fixed duration contracts are another mechanism for creating jobs, as was the case with active labor market policies, some counteracting forces may limit their effectiveness as well.

8. Conclusions

In this chapter, we have examined evidence on the impact of interventions in the labor market on wages and employment. An international comparative framework was adopted to exploit the wide variation in labor market institutions across OECD countries that have otherwise similar levels of education, living standards, and technological development. Our overall conclusion is that institutions do appear to matter. At the micro level, there was, in general, more consistently robust evidence that institutions affect the distribution of wages than employment levels.

Centralized union wage setting institutions lower wage dispersion compared to decentralized institutions, with particularly strong effects at the bottom of the distribution. Thus, the less skilled have much higher relative wages in continental Europe than in the US. Along the same lines, centralized wage setting greatly raises the ratio in pay between women and men, not only by lowering the returns to skill but also by setting minimum pay rates across diverse units. This also leads to much smaller interindustry wage differentials, again, with particularly large effects at the bottom of the distribution; the US has by far the largest interindustry wage differentials, followed by the UK and then the other countries of continental Europe. Government wage-setting interventions have also been found to have a noticeable impact on the wage distribution, including minimum wage laws and anti-discrimination policies as applied to the gender pay gap. The latter appear to have stronger immediate effects when combined with centralized wage-setting.

Evidence of the employment impact of labor market institutions is considerably more mixed. More generous severance pay mandates do appear in most studies to be associated with lower employment-to-population ratios. This indirectly implies that, given rigid wage-setting institutions, the impact of such requirements cannot be fully compensated for by adjustments in starting pay. Union policies compressing pay differentials have been found in some studies but not in others to have disemployment effects and to cause

increased reliance on public employment. In addition, more generous unemployment insurance systems appear to lead to greater levels of long-term unemployment. However, direct government intervention in pay setting through minimum wage laws or anti-discrimination policy generally has not been found to have large or in many cases even any negative employment effects. European countries have developed responses to some of the perceived adverse effects of labor market rigidities, including short-time compensation, public employment, fixed duration employment contracts, and active labor market programs. These responses may mask some of the adverse employment effects of labor market institutions which would otherwise have occurred.

To what extent are labor market institutions responsible for Europe's persistently high unemployment rate in comparison to that in the US? We first observe that there is a circumstantial case that these institutions have played a role: the US with its more flexible labor markets has low unemployment and Europe with its more rigid wage structures and greater labor market intervention has high unemployment. However, before 1973, Europe had relatively low unemployment despite its more interventionist labor market policies and more extensive collective bargaining. It is possible that in the presence of labor market rigidities, it took Europe much longer to get over the shocks of the 1970s and 1980s than the US. Real wage levels in Europe rose relative to those in the US from 1979 to 1990, and this development may have contributed to Europe's rising relative unemployment levels (Freeman, 1994). Moreover, in addition to possibly contributing to superior real wage growth, continental Europe's institutions have played a role in limiting the rise in wage inequality. Rising wage inequality has been most evident in the more laissez faire labor markets of the US and the UK. Thus, while it is easy to point to sluggish employment growth as an adverse outcome of less flexible European labor market institutions, the experience of the US shows that job creation does not come without important costs as well.

The large variation in institutions and economic performance across OECD countries raises the question of whether such differences can persist in a global economy. For example, can one country successfully mandate employment protection without pricing itself out of international markets? Will movements of capital, labor and goods undo the effects of particular government policies and institutions? While an increasingly integrated international economy can potentially place limits on what individual countries can do, Ehrenberg (1994) points out that there are at least two mechanisms that could potentially allow each country to "go its own way" in creating labor market institutions and interventions. First, many mandated benefits may be passed back to wages with little overall effect on labor costs and therefore competitiveness. Second, even in the case where institutional rigidity keeps wages from adjusting downward in the face of government mandates, exchange rates can adjust in response to changing production costs. The use of exchange rates as an adjustment outlet will become more difficult as Europe moves to a common currency. Yet we still expect that Europe's institutions will remain quite different from those in the US, giving us considerable opportunities to study the impact of these differences.

Acknowledgements

We are indebted to Orley Ashenfelter, David Card and participants at the conference for contributors to this volume, which was held at Princeton University in September 1997.

References

Abowd, J.M. and M.L. Bognanno (1995), "International differences in executive and managerial compensation", in: R. Freeman and L. Katz, eds., Differences and changes in wage structures (University of Chicago Press, Chicago, IL) pp. 67–103.

Abowd, J., F. Kramarz, T. Lemieux and D. Margolis (1999), "Minimum wages and youth unemployment in France and the US", in: D.G. Blanchflower and R.B. Freeman, eds., Youth employment and joblessness in advanced countries (University of Chicago Press, Chicago, IL) in press.

Abraham, K. and S. Houseman (1994), "Does employment protection inhibit labor market flexibility? lessons from Germany, France, and Belgium", in: R. Blank, ed., Social protection versus economic flexibility: is there a trade-off? (University of Chicago Press, Chicago, IL) pp. 59–93.

Abraham, K. and S. Houseman (1995), "Earnings inequality in Germany", in: R.B. Freeman and L.F. Katz, eds., Differences and changes in wage structures (University of Chicago Press, Chicago, IL) pp. 371–403.

Addison, J.T. and J.-L. Grosso (1996), "Job security provisions and employment: revised estimates", Industrial Relations 35: 585–603.

Agell, J. and K.E. Lommerud (1992), "Union egalitarianism as income insurance", Economica 59: 295–310.

Albæk, K., M. Arai, R. Asplund, E. Barth and S. Marsden (1996), "Inter-industry Wage differentials in the Nordic countries", in: N. Westergård-Nielsen, ed., Wage differentials in the Nordic countries (Part 1 of E. Wodensjo, ed., The Nordic Labour Markets in the 1990's) (North-Holland, Amsterdam) pp. 83–111.

Angrist, J. and A. Krueger (1991), "Does compulsory school attendance affect schooling and earnings?" Quarterly Journal of Economics 106: 979–1014.

Barth, E. and J. Zweimüller (1992), "Labour market institutions and the industry wage distribution: evidence from Austria, Norway and the U.S.", Empirica–Austrian Economic Papers 19: 181–201.

Bell, D., R. Elliott and A. Skalli (1996), "Wage structure in Britain and France: an analysis of the returns to education and age", Actes du Colloque 1er et 2me Février (INSEE, Paris) pp. 182–197.

Beller, A. (1979), "The impact of equal employment opportunity laws on the male/female earnings differential", in: C.B. Lloyd, E. Andrews and C. Gilroy, eds., Women in the labor market (Columbia University Press, New York) pp. 304–330.

Bentolila, S. and J. Dolado (1994), "Labour flexibility and wages: lessons from Spain", Economic Policy 18: 53–99.

Bertola, G. (1992), "Labor turnover costs and average labor demand", Journal of Labor Economics 10: 389–411.

Bertola, G. and A. Ichino (1995), "Crossing the river: a comparative perspective on Italian employment dynamics", Economic Policy 21: 361–420.

Björklund, A. and R.B. Freeman (1997), "Generating equality and eliminating poverty, the Swedish way", in: R.B. Freeman, R. Topel and B. Swedenborg, eds., The welfare state in transition: reforming the Swedish model (University of Chicago Press, Chicago, IL) pp. 33–78.

Blackburn, M.L. and D.E. Bloom (1993), "The distribution of family income: measuring and explaining changes in the 1980s for Canada and the United States", in: D. Card and R.B. Freeman, eds., Small differences that matter (University of Chicago Press, Chicago, IL) pp. 233–265.

Blanchard, O.J. and L.H. Summers (1986), "Hysteresis and the European unemployment problem", in: S. Fischer, ed., NBER Macroeconomics annual 1986 (MIT Press, Cambridge, MA).

Blanchflower, D. (1996), "The role and influence of trade unions in the OECD", Unpublished working paper (Dartmouth College).

Blanchflower, D. and R. Freeman (1992), "Unionism in the U.S. and other advanced O.E.C.D. countries", Industrial Relations 31: 56–79.

Blanchflower, D. and R. Freeman (1994), "Did the Thatcher reforms change British labour performance?", in: R. Barrell, ed., The UK labour market: comparative aspects and institutional developments (Cambridge University Press, Cambridge, MA) pp. 51–92.

Blank, R. (1994), "Public sector growth and labor market flexibility: the United States versus the United Kingdom", in: R. Blank, ed., Social protection versus economic flexibility: is there a trade-off? (University of Chicago Press, Chicago, IL) pp. 223–264.

Blank, R. (1997), "Is there a trade-off between unemployment and inequality? no easy answers: labor market problems in the United States versus Europe", Public policy brief no. 33 (Jerome Levy Institute, Bard College).

Blank, R. and R.B. Freeman (1994), "Evaluating the connection between social protection and economic flexibility", in: R. Blank, ed., Social protection versus economic flexibility: is there a trade-off? (University of Chicago Press, Chicago, IL) pp. 21–41.

Blau, F.D. (1977), Equal pay in the office (D.C. Heath, Lexington, MA).

Blau, F.D. and A. Beller (1988), "Trends in earnings differentials by gender, 1971–1981", Industrial & Labor Relations Review 41: 513–529.

Blau, F.D. and L.M. Kahn (1981), "Race and sex differences in quitting by young workers", Industrial & Labor Relations Review 34: 563–577.

Blau, F.D. and L.M. Kahn (1995), "The gender earnings gap: some international evidence", in: R. Freeman and L. Katz, eds., Differences and changes in wage structures (University of Chicago Press, Chicago, IL) pp. 105–143.

Blau, F.D. and L.M. Kahn (1996a), "Wage structure and gender earnings differentials: an international comparison", Economica 63 (supplement): S29–S62.

Blau, F.D. and L.M. Kahn (1996b), "International differences in male wage inequality: institutions versus market forces", Journal of Political Economy 104: 791–837.

Blau, F.D. and L.M. Kahn (1997), "Swimming upstream: trends in the gender wage differential in the 1980s", Journal of Labor Economics 15: 1–42.

Blau, F.D. and L.M. Kahn (1999), "Gender and youth employment outcomes: the US and West Germany, 1984-91", in: D.G. Blanchflower and R.B. Freeman, eds., Youth employment and joblessness in advanced countries (Univeristy of Chicago Press, Chicago, IL) in press.

Blau, F.D., M. Ferber and A. Winkler (1998), The economics of women, men and work, 3rd edition (Prentice Hall, Englewood Cliffs, NJ).

Borland, J. (1996), "Union effects on earnings dispersion in Australia, 1986–1994", British Journal of Industrial Relations 34: 237–248.

Bulow, J. and L. Summers (1986), "A theory of dual labor markets with application to industrial policy, discrimination, and Keynesian unemployment", Journal of Labor Economics 4: 376–414.

Burtless, G. (1987), "Jobless pay and high European unemployment", in: R.Z. Lawrence and C. Schultze, eds., Barriers to European growth: a transatlantic view (Brookings, Washington, DC) pp. 105–174.

Calmfors, L. (1993), "Centralisation of wage bargaining and macroeconomic performance – a survey", OECD Economic Studies 21: 161–191.

Calmfors, L. and J. Driffill (1988), "Centralization of wage bargaining", Economic Policy 3: 14–61.

Calmfors, L. and A. Forslund (1991), "Real wages and labour market policies", The Economic Journal 408: 1130–1148.

Card, D. (1996), "The effect of unions on the structure of wages: a longitudinal analysis", Econometrica 64: 957–979.

Card, D. and R.B. Freeman, eds. (1993), Small differences that matter (University of Chicago Press, Chicago, IL).

Card, D. and A.B. Krueger (1995), Myth and measurement: the new economics of the minimum wage (Princeton University Press, Princeton, NJ).

Card, D., L.F. Katz and A.B. Krueger (1994), "Comment on David Neumark and William Wascher, 'employment

effects of minimum wage and subminimum wages: panel data on state minimum wage laws'", Industrial & Labor Relations Review 48: 487–496.

Card, D., F. Kramarz and T. Lemieux (1995), "Changes in the relative structure of wages and employment: a comparison of the United States, Canada, and France", Working paper (Princeton University, Princeton, NJ).

Castillo-Freeman, A. and R.B. Freeman (1992), "When the minimum wage really bites: the effect of the U.S. level minimum on Puerto Rico", in: G.J. Borjas and R.B. Freeman, eds., Immigration and the work force (University of Chicago Press, Chicago, IL) pp. 177–211.

Chamberlain, G. (1991), "Quantile regression, censoring, and the structure of wages", Unpublished working paper (Harvard University, Harvard, MA).

Christie, V. (1992), "Union wage effects and the probability of union membership", Economic Record 68: 43–56.

Danthine, J.-P. and J. Hunt (1994), "Wage bargaining structure, employment and economic integration", The Economic Journal 104: 528–541.

Davis, S.J. and J. Haltiwanger (1991), "Wage dispersion between and within u.s. manufacturing plants", Brookings Papers on Economic Activity: Microeconomics: 115–180.

Davis, S.J. and M. Henrekson (1997), "Industry distribution of employment", Working paper no. 6246 (NBER, Cambridge, MA).

Deere, D., K.M. Murphy and F. Welch (1995), "Employment and the 1990–1991 minimum-wage hike", American Economic Review 85: 232–237.

de la Rica, S. and T. Lemieux (1994), "Does public health insurance reduce labor market flexibility or encourage the underground economy? evidence from Spain and the United States", in: R. Blank, ed., Social protection versus economic flexibility: is there a trade-off? (University of Chicago Press, Chicago, IL) pp. 265–299.

Dell'Aringa, C. and C. Lucifora (1994), "Wage dispersion and unionism: do unions protect low pay?" International Journal of Manpower 15: 150–169.

Dertouzos, J.N. and L.A. Karoly (1993), "Employment effects of worker protection: evidence from the United States", in: C. Buechtemann, ed., Employment security and labor market behavior (ILR Press, Ithaca, NY) pp. 215–227.

DiNardo, J. and T. Lemieux (1997), "Diverging male wage inequality in the United States and Canada, 1981-1988: do institutions explain the difference?" Industrial & Labor Relations Review 50: 629–651.

DiNardo, J., N.M. Fortin and T. Lemieux (1996), "Labor market institutions and the distribution of wages, 1973-1992: a semiparametric approach", Econometrica 64: 1001–1044.

Dolado, J., F. Kramarz, S. Machin, A. Manning, D. Margolis, and C. Teulings (1996), "The economic impact of minimum wages in Europe", Economic Policy 23: 319–372.

Dolton, P., D. O'Neill and O. Sweetman (1996), "Gender differences in the changing labor market", Journal of Human Resources 31: 549–565.

Edin, P.-A. (1993), "Swimming with the tide: solidarity wage policy and the gender earnings gap", Unpublished working paper (Uppsala University, Uppsala, Sweden).

Edin, P.-A. and B. Holmlund (1995), "The Swedish wage structure: the rise and fall of solidarity wage policy", in: R. Freeman and L. Katz, eds., Differences and changes in wage structures (University of Chicago Press, Chicago, IL) pp. 307–343.

Edin, P.-A. and R. Topel (1997), "Wage policy and restructuring: the Swedish labor market since 1960", in: R.B. Freeman, R. Topel and B. Swedenborg, eds., The welfare state in transition: reforming the Swedish model (University of Chicago Press, Chicago, IL) pp. 155–201.

Edin, P.-A. and J. Zetterberg (1992), "Interindustry wage differentials: evidence from Sweden and a comparison with the United States", American Economic Review 82: 1341–1349.

Ehrenberg, R.G. (1994), Labor markets and integrating national economies (Brookings, Washington, DC).

Ehrenberg, R.G. and R.S. Smith (1997), Modern labor economics, 6th edition (Addison Wesley, Reading, MA).

Ehrenberg, R.G., L. Danziger and G. San (1983), "Cost-of-living adjustment clauses in union contracts: a summary of results", Journal of Labor Economics 1: 215–245.

Erickson, C. and A. Ichino (1995), "Wage differentials in Italy: market forces and institutions", in: R. Freeman

and L. Katz, eds., Differences and changes in wage structures (University of Chicago Press, Chicago, IL) pp. 265–305.

Espinosa, M.P. and C. Rhee (1989), "Efficient Wage Bargaining as a Repeated Game", Quarterly Journal of Economics 104: 565–588.

EIRR (1992), "Minimum Pay in 18 Countries", European Industrial Relations Review 225: 14–21.

Farber, H.S. (1986), "The analysis of union behavior", in: O. Ashenfelter and R. Layard, eds., Handbook of labor economics, vol. II (North-Holland, Amsterdam) pp. 1039–1089.

Forslund, A. and A.B. Krueger (1997), "An evaluation of the Swedish active labor market policy: new and received wisdom", in: R.B. Freeman, R. Topel and B. Swedenborg, eds., The welfare state in transition: reforming the Swedish model (University of Chicago Press, Chicago, IL) pp. 267–298.

Freeman, R.B. (1980), "Unionism and the dispersion of wages", Industrial and Labor Relations Review 34: 3–23.

Freeman, R.B. (1982), "Union wage practices and wage dispersion within establishments", Industrial and Labor Relations Review 36: 3–21.

Freeman, R.B. (1988), "Labour markets", Economic Policy 3: 63–80.

Freeman, R.B. (1993), "How much has de-unionization contributed to the rise in male earnings inequality?" in: S. Danziger and P. Gottschalk, eds., Uneven tides: rising inequality in America (Russell Sage Foundation, New York) pp. 133–163.

Freeman, R.B. (1994), "How labor fares in advanced economies", in: R.B. Freeman, ed., Working under different rules (Russell Sage Foundation, New York) pp. 1–28.

Freeman, R.B. and R.S. Gibbons (1995), "Getting together and breaking apart: the decline of centralized collective bargaining", in: R.B. Freeman and L.F. Katz, eds., Differences and changes in wage structures (University of Chicago Press, Chicago, IL) pp. 345–370.

Freeman, R.B. and L.F. Katz, eds. (1995), Differences and changes in wages in wage structures (University of Chicago Press, Chicago, IL).

Freeman, R.B. and J. Pelletier (1990), "The impact of industrial relations legislation on British union density", British Journal of Industrial Relations 28: 141–164.

Friesen, J. (1996), "The response of wages to protective labor legislation: evidence from Canada", Industrial & Labor Relations Review 49: 243–255.

Gosling, A. and S. Machin (1995), "Trade unions and the dispersion of earnings in British establishments, 1980-90", Oxford Bulletin of Economics and Statistics 57: 167–184.

Green, D.A. (1991), "A comparison of estimation approaches for the union-nonunion wage differential", Department of economics discussion paper 91-13 (University of British Columbia, BC, Canada).

Gregory, R. and A. Daly (1991), "Can economic theory explain why Australian women are so well paid relative to their U.S. counterparts?" in: S.L. Willborn, ed., Women's wages: stability and changes in six industrialized countries (JAI Press, Greenwich, CT) pp. 81–125.

Gregory, R.G. and R.C. Duncan (1981), "Segmented labor market theories and the Australian experience of equal pay for women", Journal of Post Keynesian Economics 3: 403–428.

Groshen, E.L. (1991), "The structure of the female/male wage differential: is it who you are, what you do, or where you work?" Journal of Human Resources 26: 457–472.

Gruber, J. (1994), "The incidence of mandated maternity benefits", American Economic Review 84: 622–641.

Gunderson, M. and R. Robb (1991), "Equal pay for work of equal value: Canada's experience", in: D. Sockell, D. Lewin and D. Lipsky, eds., Advances in industrial and labor relations, Vol. 5 (JAI Press, Greenwich, CT) pp. 151–168.

Hamermesh, D.S. (1993), "Employment protection: theoretical implications and some U.S. evidence", in: C. Buechtemann, ed., Employment security and labor market behavior (ILR Press, Ithaca, NY) pp. 126–147.

Hashimoto, M. (1990), "Employment and wage systems in Japan and their implications for productivity", in: A.S. Blinder, ed., Paying for productivity (Brookings, Washington, DC) pp. 245–294.

Hendricks, W.E. and L.M. Kahn (1982), "The determinants of bargaining structure in U.S. manufacturing industries", Industrial and Labor Relations Review 35: 181–195.

Heylen, F. (1993), "Labour market structures, labour market policy and wage formation in the OECD", Labour 7: 25–51.

Hibbs, D.A. (1990), "Wage compression under solidarity bargaining in Sweden", Economic research report No. 30 (Trade Union Institute for Economic Research, Stockholm).

Houseman, S. (1997), "Temporary, part-time, and contract employment in the United States: new evidence from an employer survey", Unpublished manuscript (W.E. Upjohn Institute for Employment Research, Kalamazoo, MI).

Hunt, J. (1997), "The transition in East Germany: when is a ten point fall in the gender wage gap bad news?" Working paper no. 6167 (NBER, Cambridge, MA).

Juhn, C., K.M. Murphy and B. Pierce (1991), "Accounting for the slowdown in black-white wage convergence", in: M.H. Kosters, ed., Workers and their wages (AEI Press, Washington DC) pp. 107–143.

Kahn, L.M. (1993), "Unions and cooperative behavior: the effect of discounting", Journal of Labor Economics 11: 680–703.

Kahn, L.M. (1998a), "Against the wind: bargaining recentralisation and wage inequality in Norway, 1987–1991", The Economic Journal 108: 603–645.

Kahn, L.M. (1998b), "Collective bargaining and the interindustry wage structure: international evidence", Economica 65: 507–534.

Kahn, L.M. and M. Curme (1987), "Unions and nonunion wage dispersion", The Review of Economics and Statistics 69: 600–607.

Katz, H.C. (1993), "The decentralization of collective bargaining: a literature review and comparative analysis", Industrial and Labor Relations Review 47: 3–22.

Katz, L.F. and A.B. Krueger (1992), "The effect of the new minimum wage law in a low-wage labor market", Industrial & Labor Relations Review 46: 6–21.

Katz, L.F., G.W. Loveman and D.G. Blanchflower (1995), "A comparison of changes in the structure of wages in four OECD countries", in: R.B. Freeman and L.F. Katz, eds., Differences and changes in wage structures (University of Chicago Press, Chicago, IL) pp. 25–65.

Kidd, M.P. and M. Shannon (1996), "The gender wage gap: a comparison of Australia and Canada", Industrial & Labor Relations Review 49: 729–746.

Killingsworth, M. (1990), The economics of comparable worth (W.E. Upjohn Institute for Employment Research, Kalamazoo, MI).

Kornfeld, R. (1993), "The effects of union membership on wages and employee benefits: the case of Australia", Industrial & Labor Relations Review 47: 114–128.

Krueger, A.B. (1995), "The effect of the minimum wage when it really bites: a reexamination of the evidence from Puerto Rico", in: S. Polachek, ed., Research in labor economics (JAI Press, Greenwich, CT) pp. 1–22.

Krueger, A. B. and J.-S. Pischke (1997), "Observations and conjectures on the U.S. employment miracle", Working paper no. 390 (Industrial Relations Section, Princeton University, Princeton, NJ).

Krueger, A.B. and L.H. Summers (1988), "Efficiency wages and the inter-industry wage structure", Econometrica 56: 259–293.

Layard, R. and S. Nickell (1990), "Is unemployment lower if unions bargain over employment?" Quarterly Journal of Economics 105: 773–787.

Layard, R., S. Nickell and R. Jackman (1991), Unemployment (Oxford University Press, Oxford, UK).

Lazear, E.P. (1990), "Job security provisions and employment", Quarterly Journal of Economics 105: 699–726.

Lemieux, T. (1993), "Unions and wage inequality in Canada and the United States", in: D. Card and R. Freeman, eds., Small differences that matter (University of Chicago Press, Chicago, IL) pp. 69–107.

Leonard, J.S. and M. Van Audenrode (1993), "Corporatism run amok: job stability and industrial policy in Belgium and the United States", Economic Policy 8: 356–400.

Levine, D.I. and L.D. Tyson (1990), "Participation, productivity, and the firm's environment", in: A.S. Blinder, ed., Paying for productivity (Brookings, Washington, DC) pp. 183–243.

Lewis, H. G. (1986), Union relative wage effects: a survey (University of Chicago Press, Chicago, IL).

Light, A. and M. Ureta (1992), "Panel estimates of male and female job turnover behavior: can female nonquitters be identified?" Journal of Labor Economics 10: 156–181.
Lindbeck, A. and D.J. Snower (1986), "Wage setting, unemployment, and insider-outsider relations", American Economic Review 76: 235–239.
Ljungqvist, L. (1995), "Wage structure as implicit insurance on human capital in developed versus underdeveloped countries", Journal of Development Economics 46: 35–50.
Machin, S. and A. Manning (1994), "The effects of minimum wages on wage dispersion and employment: evidence from the U.K. wage councils", Industrial & Labor Relations Review 47: 319–329.
Madden, J. F. (1973), The economics of sex discrimination (DC Heath, Lexington, MA).
Main, B. (1996), "The union relative wage gap", in: D. Gallie, R. Penn and M. Rose, eds., Trade unionism in recession (Oxford University Press, Oxford, UK).
Maloney, T. (1994), "Estimating the effects of the employment contracts act on employment and wages in New Zealand", Australian Bulletin of Labour 20: 320–343.
Manning, A. (1987), "An integration of trade union models in a sequential bargaining framework", The Economic Journal: 121–139.
Manning, A. (1994), "How robust is the microeconomic theory of the trade union?" Journal of Labor Economics 12: 430–459.
Manning, A. (1996), "The equal pay act as an experiment to test theories of the labour market", Economica 63: 191–212.
Meitzen, M.E. (1986), "Differences in male and female job-quitting behavior", Journal of Labor Economics 4: 151–167.
Miller, P. and C. Mulvey (1993), "What do Australian unions do?" Economic Record 69: 315–342.
Moene, K.O. (1988), "Unions' threats and wage determination", The Economic Journal 98: 471–483.
Mulvey, C. (1986), "Wage levels: do unions make a difference?" in: J. Niland, ed., Wage fixation in Australia (Allen and Unwin, Sydney).
Neumark, D. and W. Wascher (1992), "Employment effects of minimum and subminimum wages: panel data on state minimum wage laws", Industrial & Labor Relations Review 46: 55–81.
Nickell, S. (1997), "Unemployment and labor market rigidities: Europe versus North America", Journal of Economic Perspectives: 55–74.
Nickell, S. and B. Bell (1996a), "Changes in the distribution of wages and unemployment in OECD countries", American Economic Review 86: 302–307.
Nickell, S and B. Bell (1996b), "Would cutting payroll taxes on the unskilled have a significant impact on unemployment?" Discussion paper no. 276 (London School of Economics Centre for Economic Performance, London).
OECD (1983), Employment outlook: September 1983 (OECD, Paris).
OECD (1988), Employment outlook: September 1988 (OECD, Paris).
OECD (1990), Employment outlook: July 1990 (OECD, Paris).
OECD (1993), Employment outlook: July 1993 (OECD, Paris).
OECD (1994a), Employment outlook: July 1994 (OECD, Paris).
OECD (1994b), The OECD jobs study (OECD, Paris).
OECD (1996), Employment outlook: July 1996 (OECD, Paris).
O'Neill, J. and S. Polachek (1993), "Why the gender gap in wages narrowed in the 1980s", Journal of Labor Economics 11: 205–228.
O'Neill, J., M. Brien and J. Cunningham (1989), "Effects of comparable worth policy: evidence from Washington State", American Economic Review 79: 305–309.
Pelling, Henry (1960), American labor (University of Chicago Press, Chicago, IL).
Persson, Mats (1995), "Why are taxes so high in egalitarian societies?" Scandinavian Journal of Economics 97: 569–580.
Riddell, W.C. (1993), "Unionization in Canada and the United States: a tale of two countries", in: D. Card and R.B. Freeman, eds., Small differences that matter (University of Chicago Press, Chicago, IL) pp. 109–147.

Rowthorn, R.E. (1992), "Centralisation, employment and wage dispersion", The Economic Journal 102: 506–523.

Ruhm, C.J. and J.L. Teague (1997), "Parental leave policies in Europe and North America", in: F.D. Blau and R.G. Ehrenberg, eds., Gender and family issues in the workplace (Russell Sage Foundation, New York) pp. 133–156.

Saint-Paul, G. (1996), "Exploring the political economy of labour market institutions", Economic Policy 23: 265–315.

Schmidt, C. (1995), "Relative wage effects of German unions", Unpublished working paper (University of Munich, Munich).

Schmidt, C. and K. Zimmermann (1991), "Work characteristics, firm size and wages", The Review of Economics and Statistics 73: 705–710.

Schmitt, J. (1995), "The changing structure of male earnings in Britain, 1974–1988", in: R.B. Freeman and L.F. Katz, eds., Differences and changes in wage structures (University of Chicago Press, Chicago, IL) pp. 177-204.

Siebert, H. (1997), "Labor market rigidities: at the root of unemployment in Europe", Journal of Economic Perspectives 11: 37–54.

Simpson, W. (1991), "The impact of unions on the structure of Canadian wages: an empirical analysis with microdata", Canadian Journal of Economics 18: 164–181.

Social Security Administration (1993), "Social security programs in the United States", Social Security Bulletin 56: 3–98.

Soskice, D. (1990), "Wage determination: the changing role of institutions in advanced industrialized countries", Oxford Review of Economic Policy 6: 36–61.

Summers, L.H., J. Gruber and R. Vergara (1993), "Taxation and the structure of labor markets: the case of corporatism", Quarterly Journal of Economics 108: 385–411.

Teulings, C. and J. Hartog (1998), Corporatism or competition? Labour contracts, institutions and wage structures in international comparison (Cambridge University Press, Cambridge, UK).

Van Audenrode, M.A. (1994), "Short-time compensation, job security, and employment contracts: evidence from selected OECD countries", Journal of Political Economy 102: 76–102.

Viscusi, W.K. (1980), "Sex differences in working quitting", The Review of Economics and Statistics 62: 388-398.

Wallerstein, M., M. Golden, and P. Lange (1997), "Unions, employer associations, and wage-setting institutions in northern and central Europe, 1950–1992", Industrial & Labor Relations Review 50: 379–401.

Zabalza, A. and Z. Tzannatos (1985), "The effect of Britain's anti-discriminatory legislation on relative pay and employment", The Economic Journal 95: 679–699.

Zweimüller, J. and E. Barth (1994), "Bargaining structure, wage determination, and wage dispersion in 6 OECD countries", Kyklos 47: 81–93.

Chapter 26

CHANGES IN THE WAGE STRUCTURE AND EARNINGS INEQUALITY

LAWRENCE F. KATZ*

Harvard University and NBER

DAVID H. AUTOR*

Harvard University

Contents

Abstract 1464
JEL codes 1464
1 Introduction 1464
2 Changes in the US wage structure 1467
2.1 Changes in the US wage structure, 1963–1995, March CPS data 1470
2.2 Robustness of wage structure trends across data sources 1480
2.3 Total compensation inequality versus wage inequality 1485
2.4 Observable and unobservable components of changes in wage inequality 1489
2.5 Permanent and transitory components of earnings inequality 1493
2.6 Cohort versus time effects in inequality and the returns to education 1497
2.7 Longer-term historical changes in the US wage structure 1499
3 Changes in other advanced OECD countries 1501
4 Conceptual framework: supply, demand, and institutions 1504
5 Supply and demand factors 1509
5.1 A simple supply and demand framework 1509
5.2 Some issues in supply and demand analysis 1514
5.3 Supply and demand analysis of changes in educational wage differentials 1517
5.4 Between- and within-industry shifts in relative demand 1525
5.5 Skill-biased technological change 1530
5.6 Globalization and deindustrialization 1536
5.7 Summary 1538
6 Labor market rents and labor market institutions 1540
6.1 Industry rents 1541
6.2 Unions 1542
6.3 Minimum wage 1545
6.4 The SDI model and cross-country differences in wage structure changes 1546
7 Conclusions 1547
References 1548

* The authors are grateful to David Card, Claudia Goldin, and the participants in the Handbook of Labor Economics Conference at Princeton for helpful comments. Alicia Sasser and Jessica Wolpaw provided able research assistance. Katz thanks the National Science Foundation for research support and the Russell Sage Foundation for leave support for 1997–1998.

Handbook of Labor Economics, Volume 3. Edited by O. Ashenfelter and D. Card

Abstract

This chapter presents a framework for understanding changes in the wage structure and overall earnings inequality. The framework emphasizes the role of supply and demand factors and the interaction of market forces and labor market institutions. Recent changes in the US wage structure are analyzed in detail to highlight crucial measurement issues that arise in studying wage structure changes and to illustrate the operation of the supply-demand-institution framework. The roles of skill-biased technological change, globalization forces, changes in demographics and relative skill supplies, industry labor rents, unions, and the minimum wage in the evolution of the US wage structure are examined. Recent wage structure changes are placed in a longer-term historical perspective, and differences and similarities in wage structure changes among OECD nations are assessed.

JEL codes: J0; J3

1. Introduction

Studies of the wage structure are as old as the economics profession. Adam Smith in chapter 10 of Book I of *The Wealth of Nations* provided a comprehensive and elegant analysis of the determinants of differences in wages among individuals and employments. Smith emphasized that wage differences were determined by competitive factors (compensating differentials for differences in costs of training, probability of success, steadiness of work, and other workplace amenities), differences in individual innate abilities (which he felt were relatively unimportant), and institutional (non-competitive) factors arising from the "laws of Europe" that regulated wages, restricted labor mobility, and facilitated the creation of barriers to entry. Smith noted that shifts in demand across occupations and space could generate transitory wage differentials, but that highly elastic supply responses would tend to equalize the advantages and disadvantages of different employments over the long-run in the absence of regulatory barriers to entry. The tension found in Smith's analysis between the roles of supply and demand factors and those of institutional forces in affecting wages remains through today a key theme of research on the wage structure.

Early quantitative work on the wage structure examined differences and changes in wages by occupation (Douglas, 1930; Ober, 1948) and industry (Slichter, 1950; Cullen, 1956). Douglas (1930), a pioneer in empirical studies of the wage structure, studied the evolution of the wages of white collar (managers and clerical workers) and blue collar workers in the United States from 1890 to 1926. Douglas documented a substantial decline in the wage premium to white collar work over this period (concentrated in World War I) and argued that the rapid expansion of access to public secondary education had led the growth in the supply of qualified workers to outstrip the growth in demand. Slichter (1950)

emphasized the persistence of inter-industry wage differentials and the importance of "company wage policies" as well as skill differences as explanations for the observed pattern of differentials.

The human capital revolution of the 1960s and 1970s and the increased availability of large micro datasets with information on earnings and individual characteristics shifted the emphasis to differences in wages by education and age (or potential experience). Human capital models of lifecycle earnings arising from educational and on-the-job training investments (Becker, 1962, 1993; Ben-Porath, 1967; Mincer, 1974) provide a coherent explanation of relatively timeless qualitative features of the wage structure that have been found in almost every country and data set examined (Willis, 1986): higher earnings for more-educated workers and upward sloping and concave age-earnings profiles. But the quantitative dimensions of the wage structure do differ substantially over time (as well as across countries and even regions). Tinbergen (1974, 1975) speculated that the evolution of technology tends to increase the demand for more-educated labor and characterized the evolution of the wage structure as a "race between technological development and access to education."

Research on changes in the wage structure and earnings inequality for the United States and other OECD countries has literally exploded over the past decade. The reasons for this increased research emphasis on understanding wage structure changes are clear. The wage structures of some OECD nations have changed considerably in recent decades and reasonably consistent and comparable large-scale micro datasets have become increasingly available to carefully study these issues. Educational and occupational wage differentials (especially the relative earnings of college graduates) narrowed substantially in almost all advanced nations during the 1970s. But since then divergent patterns in the evolution of the wage structure have developed. Overall wage inequality and educational wage differentials have expanded greatly in the United States and the United Kingdom since end of the 1970s. A great effort has been mounted to understand these labor market changes, in part, because widening wage structure has meant widening family income and consumption inequality and associated social problems. More modest increases in overall wage inequality and skill differentials in the 1980s and 1990s are apparent in most other OECD countries.

This chapter presents a framework for understanding wage structure changes and uses this framework to assess the determinants of recent changes in the wage structures of OECD nations. The enormous range of the existing literature motivates a sharp focus on US wage structure changes to illustrate the fruitfulness of alternative methodologies.

The overall wage distribution can be decomposed into differences in wages between groups (typically defined by skill or demographic categories) and within group wage dispersion (residual wage inequality). The basic approach utilized in this chapter links relative wage and employment changes among different demographic and skill groups to changes in both the market forces of supply and demand and to labor market institutions (e.g., unions and government mandated minimum wages). Movements in within-group inequality may also reflect market forces changing the returns to (unmeasured) skills or

directly result from changes in wage setting institutions that may serve to "standardize" wages within jobs and across firms and/or industries.

This supply-demand-institution (SDI) explanation for wage structure changes has three parts (Freeman and Katz, 1994). The first is that different demographic and skill groups are assumed to be imperfect substitutes in production. Thus shifts in the supply of and demand for labor skills can alter wage and employment outcomes. Potential important sources of shifts in the relative demand among skill groups include skill-biased technological change, non-neutral changes in other input prices or supplies (e.g., capital-skill complementarity), product market shifts, and the forces of globalization (trade and outsourcing). Sources of relative supply shifts include variation in cohort size, changes in access to education and incentives for educational investments, and immigration.

The second part is that the same underlying demand and supply shocks may have differential effects on relative wages and employment depending on differences in wage-setting and other labor market institutions. The stronger the role of wage-setting institutions and the less responsive the institutions are to changes in market forces, the more the impact is likely to fall on employment rather than on wages. Regulations governing hiring and firing as well as differences in educational and training institutions may also affect how the wage structure responds to market shifts.

Third, institutional changes themselves, such as product market deregulation and changes in the extent of unionization or degree of centralization of collective bargaining, can also alter the wage structure. A key issue in assessing the impact of institutional forces on changes in the wage structure is determining the extent to which the institutional changes are "exogenous" developments (such as changes in the political climate) or largely reflect responses to supply and demand changes.

This tension between the proper interpretation of how institutions affect wage setting has led to the development of two broad empirical approaches. The first attempts to explain actual relative wage and employment changes using a supply-demand framework and (implicitly) attributes anomalies to institutional factors or unmeasured supply and demand shifts (e.g., Katz and Murphy, 1992; Murphy and Welch, 1992; Autor et al., 1998). The second takes institutional changes as exogenous and first attempts to adjust observed wages for the impact of institutional changes and then analyzes the remaining "adjusted" wage changes using a supply and demand framework (e.g., Bound and Johnson, 1992; DiNardo et al., 1996). A key outstanding conceptual and practical issue in this second approach is how to model the impact of institutions on employment as well as wages.

The remainder of this chapter is organized as follows. Section 2 documents the changes in the US wage structure over the past three decades and places these changes into longer-term historical perspective. The US wage structure has widened along several dimensions since the late 1970s, including increases in residual wage inequality as well as wage differentials by education and experience, but differences in the time patterns of these changes suggest they partially reflect distinctive phenomena. The US data and burgeoning recent literature on US wage structure changes are used to illustrate the importance of

alternative measurement choices for inferences concerning changes in overall wage inequality and different components of the wage structure. The extent to which changes in cross-section wage inequality reflect transitory or permanent components of individual lifecycle earnings variation is also examined. Section 3 briefly summarizes recent changes in the wage distributions of other advanced nations.

Section 4 develops the SDI framework for studying wage structure changes. Section 5 examines supply and demand models of wage structure changes and assesses the importance of different supply and demand factors in recent and longer-term US wage structure changes. Section 6 examines the role of changes in labor market institutions and the incidence of labor market rents on changes in the US wage structure. The role of changes in the incidence of industry rents, the decline in unionization, and changes in the minimum wage are highlighted.

The relative earnings of more-educated workers have increased substantially in the United States since 1950 despite large increases in the relative supply of the more-educated. Rapid secular growth in the relative demand for more-skilled workers appears to be a key component of any consistent explanation for the long-run evolution of the US wage structure. Part of this relative demand shift is accounted for by observed shifts in industrial structure, most arises from within-sector skill upgrading which may reflect skill-biased technological change. Fluctuations in the educational wage differentials (e.g., the narrowing of the US college wage premium in the 1970s and its substantial widening in the 1980s) are accounted for by fluctuations in the rate of growth of college workers, institutional changes (e.g., the decline of unions in the 1980s), and possibly by some recent acceleration in the pace of demand shifts favoring the more-skilled. Section 7 summarizes the key implications for future research.

2. Changes in the US wage structure

We shall use the recent US experience to illustrate alternative approaches to measuring and explaining wage structure changes. A large and growing literature documents and attempts to explain changes in the US wage structure over the past two decades.[1] Many researchers using a variety of datasets – including both household and establishment surveys – have found that wage inequality and skill differentials in earnings increased sharply in the United States from the late 1970s to the mid-1990s. There is substantial agreement among researchers and datasets concerning some of the basic "facts" that need to be explained.

Recent changes in the US wage structure can be summarized as follows:

[1] Key studies documenting the recent evolution of the US wage distribution include Bernstein and Mishel (1997), Blackburn et al. (1990), Bound and Johnson (1992), Buchinsky (1994), Davis and Haltiwanger (1991), Freeman (1997), Gottschalk (1997), Hamermesh (1999), Juhn et al. (1993), Karoly (1993), Katz and Murphy (1992), Katz and Revenga (1989), Levy and Murnane (1992), Murphy and Welch (1992, 1997), and Pierce (1997).

1. Wage dispersion increased substantially for both men and women from the end of the 1970s to the mid-1990s. The weekly earnings of the 90th percentile worker relative to the 10th percentile worker increased by over 25% for both men and women from 1979 to 1995. The available evidence suggests earnings inequality has expanded even more dramatically if one includes the very top end (top 1%) of the distribution.[2] This pattern of rising wage inequality was *not* offset by changes in non-wage compensation favoring the low-wage workers.
2. Wage differentials by education, occupation, and age (experience) have increased. The relative earnings of college graduates and those with advanced degrees increased dramatically in the 1980s. But the gender differential declined both overall and for all age and education groups in the 1980s and 1990s.
3. Wage dispersion expanded within demographic and skill groups. The wages of individuals of the same age, education, and sex (and even those working in the same occupation and industry) were much more unequal in the mid-1990s than two decades earlier.
4. Increased cross-section earnings inequality over the past two decades has not been offset by increased year-to-year earnings mobility. Permanent and transitory components of earnings variation have risen by similar amounts (Gottschalk and Moffitt, 1994). Thus year-to-year earnings instability has also increased.
5. Since these wage structure changes have occurred in a period of rather slow mean real wage growth, the real earnings of less-educated and lower-paid workers (especially young, less-educated) males appear to be lower in the 1990s than those of analogous workers two decades earlier.[3] The employment rates of less skilled workers also appear to have fallen relative to those of more skilled workers (Juhn, 1992; Murphy and Topel, 1997; Levinson, 1998).
6. Rising earnings inequality has been the dominant contributor to a substantial increase in family income inequality both from greater dispersion in the earnings of household heads and from an increased correlation in the earnings of husbands and wives (e.g., Karoly and Burtless, 1995). Inequality of consumption expenditures also expanded from the late 1970s to the early 1990s (e.g., Cutler and Katz, 1991; US Department of Labor, 1995).

[2] For example, Hall and Liebman (1998) document that the mean (median) real total compensation of Chief Executive Officers of large, publicly-traded US corporations increased by 270% (140%) from 1982 to 1994, as compared to an increase in real average total compensation per employee for the entire economy of 7% over the same period. They also find that the mean salaries of players in Major League Baseball and the National Basketball Association increased by 207% and 378% respectively from 1982 to 1994.

[3] These conclusions about real wage growth are based on using the chain-weighted personal consumption expenditures (PCE) deflator from the National Income and Product Accounts to deflate nominal earnings measures. Readers should remember that conclusions concerning changes in real earnings are clearly sensitive to potentially large biases official price indices arising from difficulties in measuring quality change and the value of new goods (Boskin et al., 1996; Moulton, 1997). Such biases in price deflators do not affect the estimates of relative wage changes that are the focus of this chapter. Furthermore, most estimates in the literature indicate the real earnings of young, less-educated men declined from 1979 to 1995 even assuming an upward bias in the PCE deflator of 1% a year.

Thus rising US wage inequality in the 1980s and 1990s has been accompanied by large increases in wage differentials by skill group and by much greater residual inequality (within group wage dispersion). The major exception to this pattern of a widening wage structure has been the substantial narrowing of wage differentials between men and women. An important motivation for understanding these wage structure changes is that diverging US labor market outcomes appear to have translated into increased inequality in economic well-being among individuals and households from the 1970s to the mid-1990s.

Much debate exists concerning the causes of recent expansions in US wage inequality and educational wage differentials. Several prominent (and not necessarily exclusive) explanations have been offered. The first attributes wage structure changes to an increased rate of growth of the relative demand for highly educated and "more-skilled" workers driven by skill-biased technological changes, largely associated with the spread of computers and microprocessor-based technologies in the workplace (Mincer, 1991; Bound and Johnson, 1992; Berman et al., 1994; Autor et al., 1998).[4] The second explanation focuses on the role of rising globalization pressures (particularly increased trade with less-developed countries and greater foreign outsourcing) in reducing manufacturing production employment and thereby shrinking the relative demand for the less educated and leading to the loss of wage premia (rents) paid to blue collar workers in some manufacturing industries (Wood, 1994, 1995, 1998; Borjas and Ramey, 1995; Feenstra and Hanson, 1996). The third attributes rising skill differentials in the 1980s and 1990s to a slowdown in the rate of growth of the relative supply of skills because of a decline in the size of the cohorts entering the labor market and an increased rate of unskilled immigration (Katz and Murphy, 1992; Murphy and Welch, 1992; Borjas et al., 1997). A fourth explanation emphasizes changes in labor market institutions including the decline in unionization, erosion of the real and relative value of the minimum wage, and changes in wage setting norms (DiNardo et al., 1996; Freeman, 1996; Lee, 1999).

Before attempting to evaluate these alternative explanations, we need to develop a more detailed understanding of both recent and historical changes in the US wage structure and of how changes in the US compare with those in other advanced countries. We further document the evolution of the US wage structure in this section and briefly summarize changes in other countries in Section 3.

Much of our knowledge of changes in the US wage structure comes from individual level earnings data from the Current Population Survey (CPS), the basic monthly household survey that is also the source of official US unemployment and labor force data. Annual earnings data and weeks worked for the previous calendar year is collected in the

[4] A related hypothesis is motivated by the spectacular increases in earnings at the extreme top end of the distribution, the rise of within-group inequality even within detailed occupations, and by Rosen's (1981) model of the economics of superstars. This approach emphasizes how changes in technology (especially those reducing communications and transportation costs) may allow the relatively highest ability individuals to sell their services to a greatly expanded market and lead to an increased concentration of economic rewards within occupations (Frank and Cook, 1995). This hypothesis seems potentially quite relevant for performing artists and possibly many professionals, but it has yet to receive much careful empirical scrutiny to determine its broader relevance for understanding wage structure changes.

Annual Demographic Supplement to the March CPS. Public use micro data from the March CPS is available starting with March 1964 and thereby providing earnings distribution information starting in 1963. Analogous data on annual earnings and weeks for the previous calendar year is available from the Public Use Micro Samples (PUMS) of the decennial Census of Population from 1940 to 1990 (covering earnings data for 1939–1989). Data on usual weekly earnings for all wage and salary workers and the hourly wage for hourly workers is available in the May CPS from 1973 to 1978 and monthly in the Outgoing Rotation Groups (ORGs) since 1979. A robust finding of rising overall wage inequality and education/skill differentials from 1979 to the mid-1990s is apparent in the March CPS, the 1980 and 1990 Census PUMS samples, the CPS ORG samples, other household surveys, as well as some available establishment surveys.[5] But some of the nuances of the timing and patterns of changes in the wage structure (especially patterns of changes in within-group or residual inequality) are somewhat sensitive to choice of data set and the precise sample and earnings concept used.

This section first summarizes changes in the US wage structure from 1963 to 1995 using data from the March CPSs. The robustness of these findings across datasets and to alternative measurement decisions is then explored. The recent changes are also compared to longer-term historical trends and used to illustrate alternative approaches to decomposing changes in the wage structure (between-group versus within-group components, permanent versus transitory components or earnings variation, and changes in "quality" between cohorts versus changes in skill prices within cohorts).

2.1. Changes in the US wage structure, 1963–1995, March CPS data

Changes in the US wage structure over the past several decades are illustrated using data on the weekly earnings of full-time, full-year, wage and salary workers (those working 35 h or more per week and working at least 40 weeks in the previous calendar year) from the

[5] Analyses of wage inequality trends using these other household surveys – the Survey of Income and Program Participation (SIPP), the National Longitudinal Survey of Youth (NLSY), and the Panel Study of Income Dynamics (PSID) – include Bernstein and Mishel (1997), Buchinsky and Hunt (1996), Gottschalk and Moffitt (1992, 1994, 1998), Haider (1997), and Lerman (1997). Studies using establishment-level datasets include Davis and Haltiwanger (1991), Dunne et al. (1997), Groshen and Levine (1997), and Pierce (1997).

[6] Information on weeks worked and usual weekly hours in the previous calendar year is available in the March CPS starting in 1976 (providing data for 1975); the earlier March CPSs only provided bracketed weeks worked information and hours worked last week. A full-time/part-time work indicator for the previous year is consistently available in all years of the March CPS public use samples. Comparisons of features of the distribution of annual or weekly earnings for full-time, full-year workers can be made rather consistently since 1963, but analyses of hourly wages or of broader sets of workers are much more consistent with a focus on data since 1975. The Census PUMSs prior to 1980 have similar limitations and do not contain a measure of usual weeks worked in the previous year. Alternative approaches to imputing hours worked in the previous calendar year in the early March CPS and Census PUMS samples are discussed in Autor et al. (1998), Juhn et al. (1993), Katz and Murphy (1992), and Murphy and Welch (1992). The basic broad patterns of changes in hourly wage distributions for full-time workers or all workers using these imputation techniques prior to 1975 are similar to those of weekly wages of full-time, full-year workers.

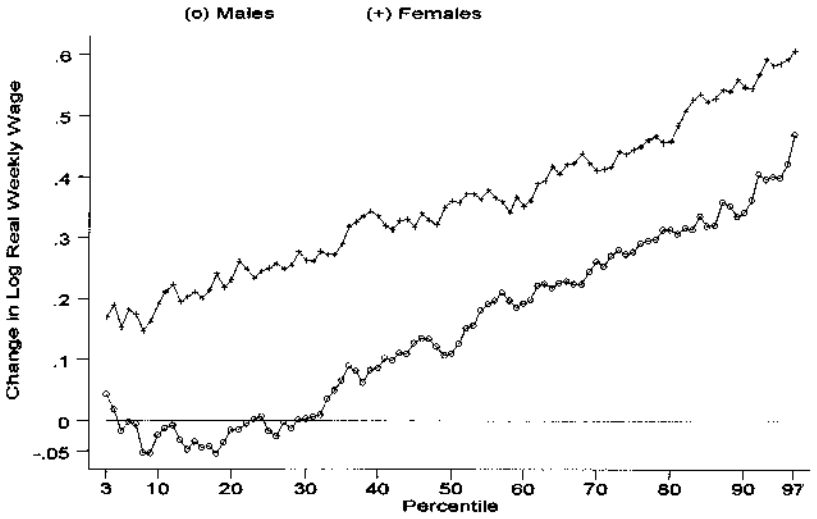

Fig. 1. Change in log real weekly wage by percentile, 1963–1995.

March CPSs of 1964 to 1996 (covering earnings from 1963 to 1995).[6] The core sample is further restricted to adults prior to retirement age (those aged 19–65 at the survey date), without allocated earnings, who earned at least $67 per week in 1982 dollars (equal to one-half of the 1982 real minimum wage based on a 40 h week).[7] Weekly earnings are imputed for those with top-coded earnings by multiplying value of the top code by 1.5. The qualitative aspects of the findings are not very sensitive to these restrictions and imputations with the exception of the treatment of outliers with extremely low weekly earnings. When workers with extremely low reported weekly earnings are kept in the sample, we find a pronounced (and implausibly large) reduction in most measures of inequality (especially for women) in the 1960s.[8] The findings reported in this section are quite similar to those of other analyses of the March CPS data including Gottschalk (1997), Juhn et al. (1993), Karoly (1993), Katz and Murphy (1992), and Murphy and Welch (1992, 1997).

Fig. 1 (following the approach of Juhn et al., 1993) plots the change in log real wages by percentile for both men and women from 1963 to 1995. The figure displays a substantial widening of both the male and female wage distributions with the wages of workers in the upper end (the 90th percentile) rising by approximately 40% (34 log points) relative to those in the lower end (the 10th percentile) for both men and women.[9] There is essentially no real wage gain from 1963 to 1995 for men in the bottom quarter of the distribution. The

[7] Nominal wages are converted into constant dollars using PCE deflator.

[8] Juhn et al. (1993) reach similar conclusions concerning the sensitivity of conclusions about inequality trends for men to alternative measurement and sample choice decisions using the March CPS data.

[9] The convention used in this chapter is to refer to log changes multiplied by 100 as changes in log points.

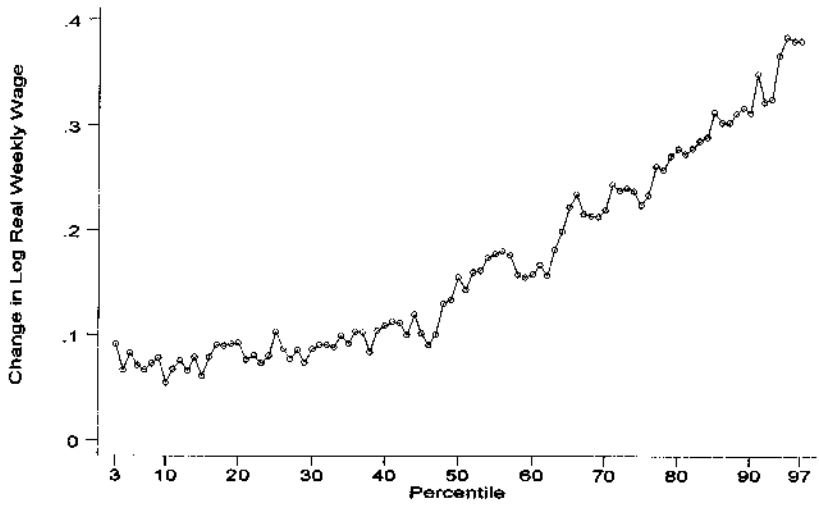

Fig. 2. Change in log real weekly wage by percentile, all, 1963–1995.

divergence of earnings is not limited to comparisons of workers at the top and the bottom. The figure indicates an almost linear spreading out of the entire wage distribution for women and for the wage distribution above the 30th percentile for men. Fig. 1 also shows that women gained on men throughout the wage distribution with the earnings of the median woman rising 27% (23 log points) relative to the median man from 1963 to 1995. Fig. 2 illustrates that the overall wage distribution (men and women combined) also spread out substantially over the past few decades, especially in the top half of the distribution.

The four panels of Fig. 3 decompose changes in wage inequality (and real earnings) from 1963 to 1995 for men and women into 4 sub-periods (1963–1971, 1971–1979, 1979–1987, and 1987–1995) that roughly correspond to the 1960s, 1970s, 1980s, and 1990s. There are some striking differences across the sub-periods. There is little overall change in wage inequality and rapid real wage growth for both men and women in the 1960s. Real wage growth slows down in the 1970s and some widening begins in the bottom half of the distribution for males. There is essentially no change in the gender gap from 1963 to 1979. The rise in wage inequality for both men and women over the entire 1963–1995 period is dominated by the rapid spreading out of the male and female wage distributions from 1979 to 1987. This pattern of rising inequality continues in a more modest form for 1987–1995. Similarly the gender gap narrows in the 1980s and 1990s.

Fig. 4 gives a sense of the full time series of changes in inequality for men and women by plotting the 90–10 log wage differential by sex annually from 1963 to 1995. Table I summarizes alternative measures of wage inequality for all, men, and women for selected years from 1963 to 1995. The Gini coefficient, standard deviation of log wages, and 90–10

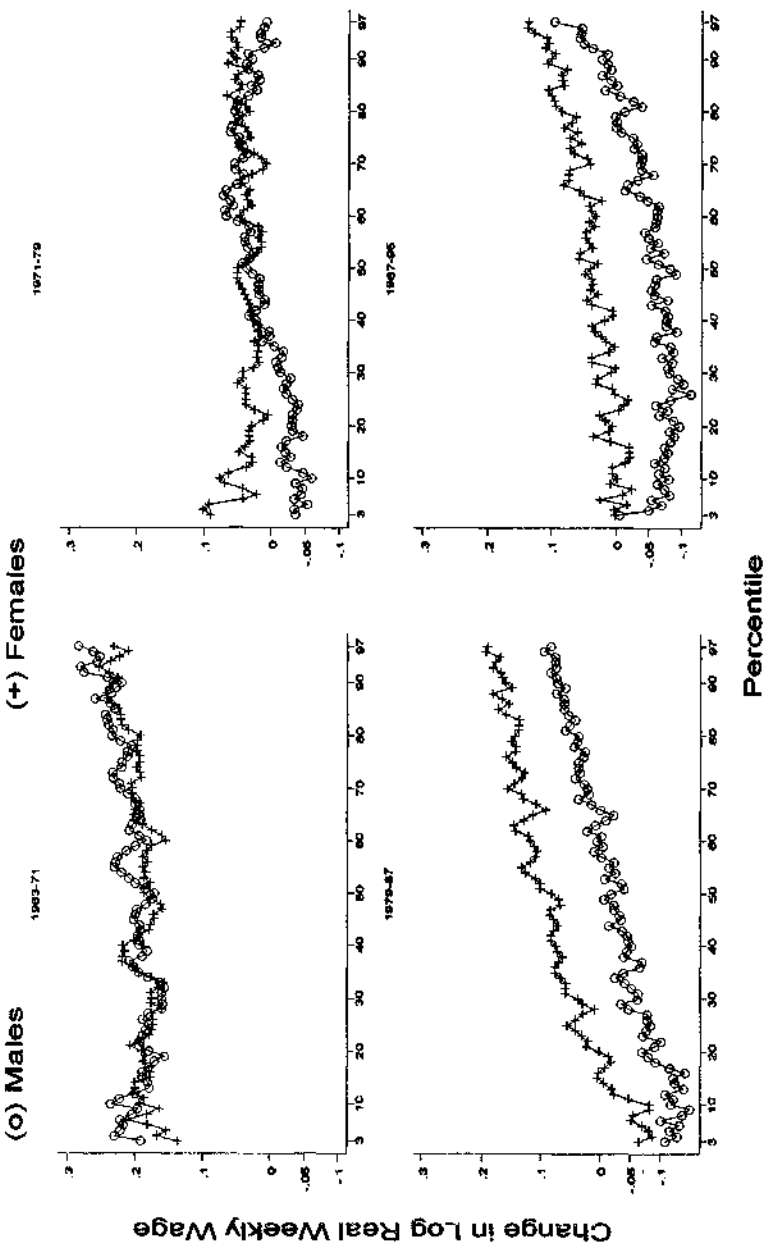

Fig. 3. Change in log real weekly wage by percentile.

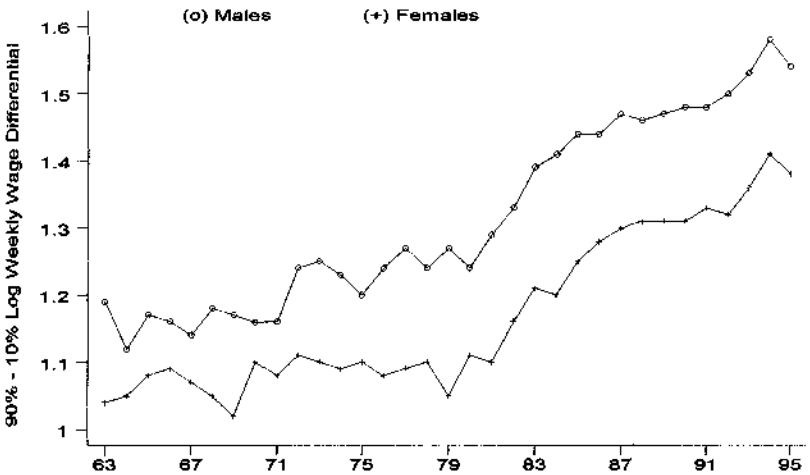

Fig. 4. Overall US wage inequality, 1963–1995.

log wage differential show somewhat similar patterns of increases in inequality for all, men, and women. The standard deviation of log wages is a useful summary measure of wage dispersion if wages are approximately log normal, but is much more sensitive to extreme outliers at the top and the bottom than are the reported quantile measures of wage dispersion. The Gini coefficient is quite sensitive to shifts in earnings in the middle of the distribution. Rising wage inequality has occurred in both the top and bottom halves of the wage distributions.

The changes in overall earnings inequality summarized in Figs. 1–4 and Table 1 reflect changes in wage differentials between demographic/skill groups and changes in inequality within groups. Table 2 summarizes the between-group changes by presenting log real wage changes from 1963 to 1995 for various groups defined by education, potential experience (age), and sex.[10] Mean (predicted) log real weekly earnings were computed in each year for 64 detailed sex-education-experience groups and mean wages for broader groups in each year are weighted averages of the relevant sub-group means using a fixed set weights (the 1980 share of total hours worked from the 1980 Census PUMS sample) to

[10] Important changes in wage differentials by race, ethnicity, and immigrant status have also occurred over the past several decades. In particular, the black/white wage differential narrowed substantially from the mid-1960s to the mid-1970s, but shows little change over the past two decades and some erosion of progress for young workers (e.g., Heckman and Donahue (1991). These dimensions of wage structure changes are beyond the scope of this chapter. See the chapter in this volume by Altonji and Blank on racial wage differentials and the chapter by Borjas on relative wage movements by immigration status.

Table 1
Measures of wage inequality for weekly wages of full-time, full-year workers, March CPS, 1963–1995

	SD of log wage	Percentiles of log wage distribution			Gini coefficient
		90–10	90–50	50–10	
A. Males					
1963	0.469	1.19	0.51	0.68	0.250
1971	0.495	1.16	0.55	0.61	0.270
1979	0.517	1.27	0.55	0.72	0.277
1987	0.579	1.47	0.65	0.82	0.313
1995	0.616	1.54	0.74	0.79	0.343
B. Females					
1963	0.406	1.04	0.50	0.54	0.223
1971	0.430	1.08	0.54	0.55	0.238
1979	0.432	1.05	0.54	0.51	0.243
1987	0.506	1.30	0.61	0.69	0.281
1995	0.544	1.38	0.68	0.70	0.304
C. All males and females					
1963	0.502	1.27	0.57	0.70	0.272
1971	0.530	1.31	0.62	0.68	0.293
1979	0.539	1.35	0.66	0.69	0.299
1987	0.580	1.44	0.70	0.74	0.320
1995	0.603	1.54	0.76	0.78	0.340

adjust for compositional changes within these broader groups.[11] The first row of Table 2 indicates that (composition-adjusted) real wages grew by 7% (or 6.6 log points) over the entire period, but this growth reflects rapid growth in the 1960s and modest declines since the early 1970s. This measure of real wage growth differs from standard measures in being

[11] The 64 sex-education-experience groups are based on a breakdown of the data into 2 sexes, 8 education categories (0–8, 9, 10, 11, 12, 13–15, 16–17, and 18+ years), and 4 potential experience categories (1–10, 11–20, 21–30, and 31+ years). Changes in the coding of education in the CPS starting in 1992 make it difficult to be fully consistent over time in defining education groups. We follow the approach suggested by Jaeger (1997a) in forming "consistent" education categories before and after the data changes. To make sure changes from 1987 to 1995 are not driven by changes in the education codes, the wage change for each group from 1990 to 1991 is calculated for full-time workers using the CPS Outgoing Rotation Groups which use the old education codes for each of these years and the 1987 to 1995 March CPS change is adjusted for the difference between the CPS ORG and March CPS change from 1990 to 1991. Log weekly wages of full-time, full-year workers are regressed each year separately by sex on the dummy variables for the 8 consistent education categories, a quartic in experience, 3 region dummies, black and other race dummies, and interactions of the experience quartic with 3 broad education categories (high school graduate, some college, and college plus). The (composition-adjusted) mean log wage for each of the 64 groups in a given year is the predicted log wage from these regressions evaluated for whites, living in the mean region based on the 1980 Census distribution of employment, at the relevant experience level (5, 15, 25 or 35 years depending on the experience group). Potential experience in the earnings year (previous calendar year) is measured as survey data age minus years of schooling minus 7.

Table 2
US real weekly wage changes for full-time, full-year workers, March CPS, 1963–1995[a]

Group	Change in mean log real weekly wage(multiplied by 100)				
	1963–1971	1971–1979	1979–1987	1987–1995	1963–1995
All	19.1	−1.4	−4.0	−7.2	6.6
Sex					
Men	20.4	−2.1	−7.3	−10.1	0.9
Women	16.9	−0.1	1.5	−2.5	15.8
Education (years of schooling):					
0–11	15.6	1.6	−10.8	−9.4	−4.5
12	17.5	1.3	−6.3	−7.1	5.5
13–15	18.6	−1.9	−2.2	−10.2	4.4
16+	26.0	−7.1	5.3	−1.8	22.4
16–17	23.0	−7.4	3.9	−2.9	16.6
18+	32.3	−6.5	8.1	5.9	34.5
Experience (men)					
5 years	19.9	−5.8	−9.7	−9.7	−5.3
25–35 years	20.1	1.4	−4.7	−10.5	6.4
Education and experience					
Education 12					
Experience 5	19.1	−0.8	−18.3	−10.7	−10.7
Experience 25–35	16.8	4.5	−4.6	−6.6	10.1
Education 16+					
Experience 5	24.2	−12.7	7.8	−8.0	11.2
Experience 25-35	34.8	−0.3	3.5	−2.0	32.9

[a] Notes: The numbers in the table represent changes in the (composition-adjusted) mean log wage for each group, using data on full-time, full-year workers from the March CPS covering calendar years 1963–1995. The data were sorted into sex-education-experience groups based on a breakdown of the data into 2 sexes, 8 education categories (0–8, 9, 10, 11, 12, 13–15, 16–17, and 18+ years), and 4 potential experience categories (1–10, 11–20, 21–30, and 31+ years). Log weekly wages of full-time, full-year workers were regressed in each year separately by sex on the dummy variables for the 8 education categories, a quartic in experience, 3 region dummies, black and other race dummies, and interactions of the experience quartic with 3 broad education categories (high school graduate, some college, and college plus). The (composition-adjusted) mean log wage for each of the 64 groups in a given years is the predicted log wage from these regressions evaluated for whites, living in the mean region based on the 1980 Census distribution of employment, at the relevant experience level (5, 15, 25 or 35 years depending on the experience group). Mean log wages for broader groups in each year represent weighted averages of the relevant (composition-adjusted) cell means using a fixed set of weights (the 1980 share of total hours worked from the 1980 Census PUMS). All earnings numbers are deflated by the chain-weighted (implicit) price deflator for personal consumption expenditures.

a geometric (rather than arithmetic) mean and by reflecting wages for a fixed demographic distribution. Hence it does not reflect changes in the level of wages arising from shifts in the education, gender, or experience composition of the work force.

The next two rows of Table 2 indicate that the (fixed-weight) mean log wage of women increased by 15 log points relative to men from 1963 to 1995 with the improvement almost

entirely concentrated in the 1980s and 1990s.[12] In fact, the earnings of women increased relative to those of men in almost all education-experience categories from 1979 to 1995. Panel A of Fig. 5 illustrates the similar time pattern of changes in the female/male log wage differential for high school graduates (those with 12 years of schooling) and college graduates (those with 16 or more years of schooling).

The next six rows of Table 2 show the evolution of real wages by education group. The real wage changes are, for the most part, increasing by education group over the full period reflecting a rise in education-based wage differentials (particularly a sharp increase in the relative earnings of those with at least a college degree). The changes in educational wage differentials differ substantially across sub-periods. College graduates (particularly those with 18 or more years of schooling) gained substantially in the 1960s, but the college wage premium narrowed (especially for younger workers in the 1970s). Educational wage differentials increased sharply from 1979 to 1987 with the college plus/high school wage differential rising by 12 log points. The relative earnings of college graduates continued rising into the 1990s, but those with some college have done particularly poorly in the 1990s. The much studied time pattern of the overall college/high school wage differential and the college/high school wage differential for young workers (those with 5 years of schooling) are shown in panel B of Fig. 5. Occupational wage differentials (e.g., the earnings of professional and managerial workers relative to production workers) also narrowed in the 1970s and then exploded in the 1980s (Blackburn et al., 1990; Murphy and Welch, 1993a).

The bottom rows of Table 2 summarize changes in real wages for older versus younger males both overall and for high school and college graduates separately. Over the entire sample period, the wage gap between older and younger males expanded with the earnings of peak earners, those with 25–35 years of experience, rising by 12 log points relative to younger workers with 5 years of experience. The differences in time pattern of changes in experience differentials for high school and college graduates are shown in panel C of Fig. 5. Experience differentials rose more sharply for college graduates in the 1960s and 1970s, then increased rapidly in the early 1980s for high school graduates and narrowed in the 1980s for college graduates. The overall change for both high school and college graduates involved substantial increases in the relative earnings of peak earners to young workers. Wage differences by age (potential experience) also expanded for women in the 1980s (Katz et al., 1995; Gottschalk, 1997).

We have so far considered wage differentials for groups distinguished by sex, education, and age/experience. But these factors account for only about one third of overall wage

[12] Real wage growth from 1963 to 1995 for both men and women is much more rapid when one uses the simple (unweighted) average weekly wage of full-time, full-year workers, rather than the fixed-weighted averages presented in Table 2. We find the unweighted average of log weekly wages increased by 0.36 for women and 0.16 for men from 1963 to 1995. Educational upgrading (rather than changes in the age distribution of workers) largely accounts for the faster growth in simple average wages than in fixed-weighted averages holding the education-experience composition of the workforce constant. Murphy and Welch (1992) report similar results for different measures of real wage growth for males from 1963 to 1989.

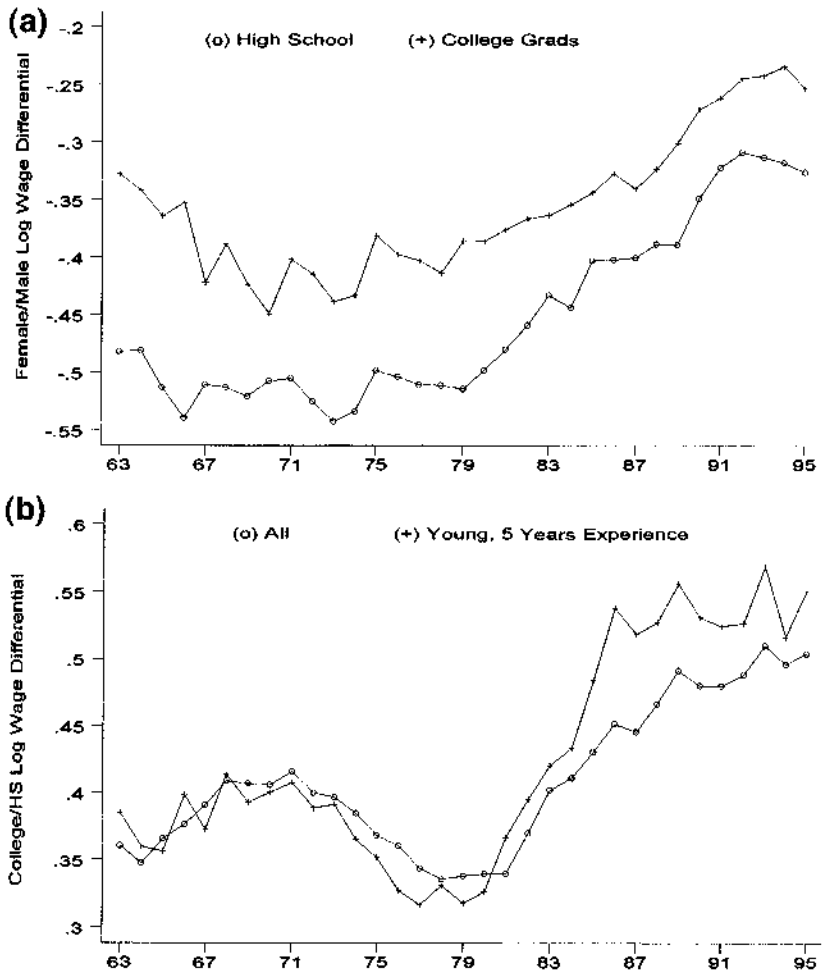

Fig. 5. (A) Female/male log weekly wage differential, 1963–1995; (B) college/HS log weekly wage differential, 1963–1995; (C) returns to experience, males, 1963–1995; (D) residual wage inequality, 90–10 differential, 1963–1995.

variation so that changes in wage dispersion within these groups are likely to be an important part of changes in the overall wage inequality. Residual (or within-group) inequality is examined here by looking at changes in the distribution of log wage residuals from separate regressions by sex each year of log weekly wages on a full set of 8 education

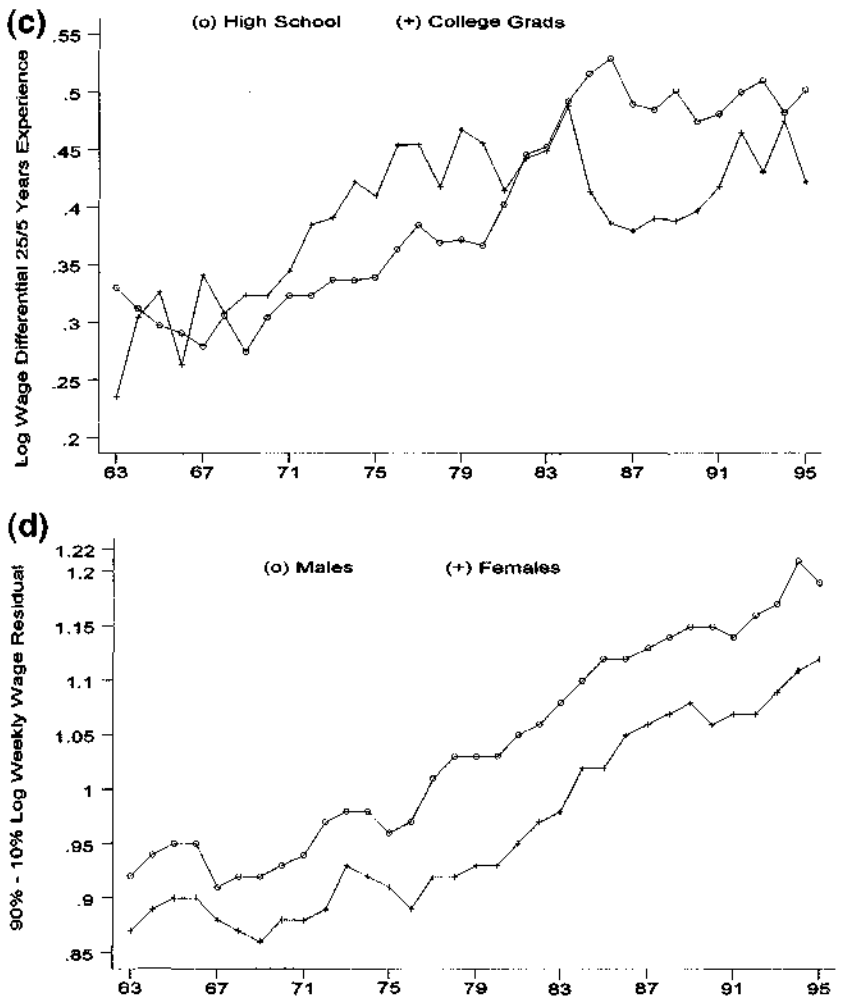

Fig. 5. *Continued*

dummies, a quartic in experience, interactions of the experience quartic with 3 broad education categories, 3 region dummies, and 2 race dummies. Panel D of Fig. 5 and Table 3 summarize the time pattern of changes in the log wage differential between the 90th and 10th percentiles of the residual wage distribution. Residual log weekly wage inequality for full-time, full-year workers increased substantially by 27 log points for men and 25 log points for women from 1963 to 1995. Residual wage inequality started increasing in the 1970s and continued rising rapidly in the 1980s and at a somewhat slower pace

in the 1990s. The rise in wage inequality within groups suggests that the "least-skilled" or least-lucky" workers within each category as well as less-educated and less-experienced workers have seen their relative earnings decline substantially over the past two decades. But the time patterns of changes in within group inequality, educational wage differentials, and experience differentials are distinctive.

In summary, we conclude from the March CPS data on the weekly wages of full-time, full-year (FTFY) workers that overall US wage inequality for both men and women expanded from the early 1960s to the mid-1990s, with changes in the 1980s accounting for much of the increase. Between- and within-group inequality increases both contributed to rising wage dispersion. More specifically, the college wage premium rose from 1963 to 1971, declined substantially in the 1970s, increased sharply in the 1980s, and continued to rise at a more modest pace in the first half of the 1990s. Experience differentials also expanded from 1963 to 1995. Relative earnings declines for young workers are largest in the 1970s for college workers and in the 1980s for the less educated. Residual wage inequality is rather stable in the 1960s, starts to increase for men in the 1970s, and increases dramatically for men and women from 1980 to 1995. After remaining fairly stable in the 1960s and 1970s, male/female wage differentials narrowed substantially in the 1980s and 1990s. The narrowing of the gender gap in earnings means that overall wage inequality for men and women combined increased by much less than wage inequality for either men or women analyzed separately. The 90–10 log weekly wage differential for all FTFY workers increased by 19 log points from 1979 to 1995 as compared to increasing by 27 log points for men and 31 log points for women over the same period.

Changes in the US wage structure over the past several decades seem, at least superficially, consistent with a general rise in the labor market returns to "skill." The returns to observed skill proxies (education, occupation, and experience) have increased, and some interpret the rise in within group inequality as reflecting a rise in the returns to unobserved skills (Juhn et al., 1993). An increase in the gap between the rate of growth of the relative demand for more-skilled workers and the relative supply of such workers represents a potential market-driven explanation for rising skill returns. The substantial decline in the gender gap since 1979 might reflect increased relative skills (e.g., actual experience and training) within education-age groups or shifts in labor demand favoring more female-intensive labor market segments (industries, occupations, particular skills). An alternative interpretation for the widening between and within group inequality is a weakening of labor market institutions and norms that compressed wages both across and within skill groups.

2.2. Robustness of wage structure trends across data sources

The basic pattern of wage structure changes from the early 1960s to the mid-1990s documented in this section for the weekly wages of FTFY workers appears rather robust and is consistent with other studies using data on weekly and hourly wages for samples from the March CPSs, Census PUMS, and the CPS May samples and ORGs (e.g., Bound

and Johnson, 1992; Katz and Murphy, 1992; Juhn et al., 1993; DiNardo et al., 1996; Bernstein and Mishel, 1997; Gottschalk, 1997; Bernard and Jensen, 1998). While we focus on the March CPS in this chapter because it provides the longest consistent US earnings series collected at high frequency, we briefly compare trends in inequality measures in the March CPS with other US data sources below.

2.2.1. *Educational differentials*

Table 3 provides comparisons of annualized log changes in the college-plus/high school, some college/high school and high school/ninth-grade wage differentials for weekly and hourly earnings for the years 1959–1996 using (as available) data from the March CPS, May CPS, CPS Outgoing Rotation Groups, and Census PUMS.[13] All samples include wage and salary workers aged 19–65 and exclude allocated observations, the lowest one-percent of earners, and those whose hourly wage exceeds the top-coded value for full-time earners.[14] Hourly samples include both full- and part-time workers while weekly earnings samples are limited to full-time workers and, in the March CPS and Census PUMS, those working 40-plus weeks. Sample weights are used throughout and are multiplied by weekly hours in hourly wage samples to weight equally all hours of labor input (e.g., DiNardo et al., 1996; Lerman, 1997).[15] Earnings are imputed for top-coded observations by multiplying the value of the top code by 1.5.

For the 1960–1996 period, trends in educational differentials are highly comparable across data sources and weekly and hourly samples and are consistent with widely documented findings. Earnings differentials expand modestly in the 1960s, contract substantially in the 1970s, expand even more dramatically during the 1980s, and continue to grow at a slower rate in the 1990s.[16] Two sources of uncertainty are worth noting. First, in the 1960s, the March CPS data indicate substantially more growth in the college-plus/high school differential than the Census PUMS, a pattern driven by very large estimated wage differentials in the March 1963 CPS.[17] Second, due to incompatibilities introduced in the CPS education measure in 1992 and the subsequent redesign of the CPS survey in 1994, estimated trends in inequality metrics are less reliable in the 1990s than in other periods.[18]

[13] The March CPS sample covers 1963–1995, the May CPS sample covers 1973–1979, the ORG sample covers 1979–1996, and the Census PUMS covers 1959–1989. All estimates of changes in wage differentials are calculated as 10 times annualized log changes to facilitate comparisons among data sources that may only be available for part of a decade (e.g., the March sample for 1963–1969). Wage differentials are estimated from separate cross sectional log earnings regressions in each year by gender and with genders combined. See the table note for further details.

[14] As noted, March samples exclude those earning less than half the 1982 minimum wage in real dollars. Allocation flags are not available for May CPS samples and hence allocated observations are retained.

[15] Census samples are weighted by weeks worked in the previous year rather than hours in the previous week.

[16] Implausibly large growth in the high-school/9th grade differential during the 1990s is most likely due to changes to the education question after 1992.

[17] As noted previously, the March data for the 1960s are quite sensitive to the treatment of extremely low hourly earnings.

[18] See Jaeger (1997a,b), Polivka (1996), Mishel et al. (1997b), and Lerman (1997) for discussion.

Table 3
Comparison of annualized changes in educational wage differentials among CPS and Census data sources: 1959–1996 (10 × annualized log changes)[a]

	Males			Females			Males & Females		
	C+/HS	Some C/ HS	HS/9th grade	C+/HS	Some C/ HS	HS/9th garde	C+/HS	Some C/ HS	HS/9th grade
A. Full-time weekly earnings									
1960s									
March CPS	0.149	−0.006	0.032	0.047	0.069	0.020	0.127	0.015	0.027
Census PUMS	0.056	−0.003	0.018	0.071	0.021	0.008	0.063	0.005	0.018
1970s									
March CPS	−0.061	−0.029	0.043	−0.028	−0.028	−0.015	−0.053	−0.026	0.022
Census PUMS	−0.027	−0.017	0.029	−0.046	0.007	−0.014	−0.026	−0.008	0.016
May CPS	−0.073	−0.007	0.004	−0.144	−0.021	−0.050	−0.102	−0.012	−0.016
1980s									
March CPS	0.167	0.065	0.068	0.139	0.064	0.083	0.158	0.065	0.072
Census PUMS	0.159	0.060	0.032	0.159	0.064	0.035	0.159	0.060	0.033
CPS ORG	0.168	0.067	0.042	0.133	0.075	0.098	0.156	0.070	0.061
1990s									
March CPS	0.142	0.039	0.038	0.088	0.035	0.089	0.120	0.037	0.057
CPS ORG	0.109	−0.002	0.161	0.138	−0.012	0.056	0.122	−0.007	0.126

B. All hourly earnings (weighted by hours worked)

1960s									
March CPS	0.116	−0.018	0.016	0.003	0.072	0.045	0.090	0.010	0.020
Census PUMS	0.063	−0.011	0.017	0.069	0.024	−0.011	0.071	0.001	0.008
1970s									
March CPS	−0.059	−0.032	0.007	−0.038	−0.008	−0.021	−0.045	−0.018	−0.007
Census PUMS	−0.039	−0.019	0.024	−0.105	−0.003	−0.021	−0.054	−0.011	0.009
May CPS	−0.055	0.001	−0.016	−0.119	−0.022	−0.076	−0.079	−0.007	−0.040
1980s									
March CPS	0.166	0.065	0.065	0.117	0.044	0.080	0.150	0.056	0.072
Census PUMS	0.144	0.054	0.035	0.135	0.059	0.027	0.144	0.057	0.031
CPS ORG	0.151	0.063	0.028	0.134	0.076	0.081	0.147	0.069	0.046
1990s									
March CPS	0.076	0.002	0.060	0.095	0.028	0.015	0.086	0.014	0.043
CPS ORG	0.093	−0.012	0.142	0.113	−0.015	0.055	0.102	−0.014	0.111

[a] Numbers above are 10 × annualized log changes in estimated log earnings differentials. Samples are: 1960s: March 1963--1969 CPS and Census PUMS 1959--1969; 1970s: March 1969--1979 CPS, Census PUMS 1969--1979, and May 1973--1979 CPS; 1980s: March 1969--1979 CPS, Census PUMs 1979--1989, and CPS ORG 1979--1989; 1990s: March 1989--1995 CPS and CPS ORG 1989--1996. All wage differentials are estimated using separate cross-sectional log earnings regressions in each sample and year that include 10 education category dummies corresponding to years of school or highest degree completed (0, 1–4, 5–6, 7–8, 9, 10, 11, some college, college graduate, post-college), a quartic in potential experience, a non-white dummy, a part-time dummy (if applicable), and three region dummies. Pooled-gender earnings regressions include a female dummy, and interactions between female and the experience quartic, part-time dummy, non-white dummy, and region dummies. All samples exclude allocated observations (except the May CPS), those whose earnings are below the lowest 1% of earners in the full sample, and those whose hourly wage exceeds the top-coded value for full-time earners. March samples are limited to those earning at least 1/2 the real value of the 1982 minimum wage converted from nominal dollars using the PCE deflator. Hourly samples include both full- and part-time workers. Weekly earnings samples are limited to full-time workers and, in the March CPS and Census PUMS, those working 40-plus weeks. Sample weights are used in all estimates and are multiplied by weekly hours in hourly wage samples or, in Census hourly samples, by weeks worked in the previous year. Earnings are imputed for top-coded observations by multiplying the value of the top code by 1.5. The college-plus/high school differential is a weighted average of the exactly college/high school differential and the post-college/high school differential where the weights are the relative employment shares of those with exactly a college education and those with post-college education from the 1980 Census PUMS (for Census samples) and the 1980 CPS ORG (for CPS samples).

To explore the robustness of these relationships, we have employed a variety of earnings cutoffs (half the minimum wage, one-third the minimum wage, 2% dropped, $0.50–250 real hourly earnings) and sub-samples (white, non-agricultural, white and non-agricultural). The time pattern of results in Table 3 is relatively insensitive to these manipulations.

2.2.2. Overall and residual earnings inequality

In contrast to our findings on educational ratios, trends in overall and residual inequality as measured by wage quantiles, the Gini coefficient, and the variance of log earnings are less consistent across data sources and are more sensitive to the choice of lower cut-off (i.e., handling of outliers), top-coding, and choice of sample (full-time, all), earnings concept (weekly, hourly) and weights (bodies, weeks, labor hours supplied).

Table 4 presents measures of annualized decadal changes in overall and residual inequality for the 1959–1996 period using the CPS and Census samples as above. The Census PUMS indicates modest expansion in overall weekly earnings inequality in the 1960s for men and women separately and combined, the bulk of which is accounted for by growth in the 90–50 log earnings ratio. Hourly earnings inequality for women, however, shows no overall increase during this period and the female hourly 50–10 ratio contracts slightly. The March CPS data for the 1960s shows slight overall contraction in inequality for both weekly and hourly samples, a pattern that is again likely to be driven by very low earnings values in the 1963 data.

The 1970s data present a largely consistent picture of stable between group inequality and growing residual inequality. Both March CPS and Census PUMS indicate moderate growth in overall male earnings inequality for both weekly and hourly earnings concentrated in the lower half of the distribution and almost entirely accounted for by the growth in the residual. Trends in male earnings inequality in the May CPS are comparable, with the exception that the May data show no growth in overall male weekly earnings inequality as measured by the 90–10 ratio. All data sources indicate either no growth or modest contraction of female earnings inequality (overall and residual) during the 1970s, with a more pronounced contraction visible in hourly samples.

Overall inequality expands dramatically across all data sources and sub-samples in the 1980s, with the expansion roughly evenly split between the upper and lower halves of the distribution for male and pooled-gender samples, and concentrated in the lower half for female samples. Trends in residual inequality are less consistent across data sources, however. While residual inequality growth accounts for approximately two-thirds of overall inequality growth in weekly and hourly samples in March and ORG CPS data during the 1980s, this is not true for the Census PUMS where the variance of log wage residuals is essentially static between 1979 and 1989 (the 90–10 residual earnings ratio in the Census indicates modest growth during this period, however).

An important pattern not visible from Table 4 is that the expansion of earnings inequality during the 1980s is not smooth but rather is concentrated in the 1979–1985 period, particularly for pooled-gender and male samples. In the ORG and March data, approxi-

mately 80% of the growth of overall male inequality, and 90% of the growth of pooled-gender inequality, occurs between 1979–1985. Residual inequality grows somewhat more smoothly during the entire decade, however, and shows little deceleration for women after 1985, especially in the March data.

The recent redesign of the CPS means trends in wage inequality during the 1990s are less certain and a subject of current debate (e.g., Bernstein and Mishel, 1997; Lerman, 1997). Our reading of the data is that overall and residual inequality in the upper half of the distribution continued to expand modestly during 1989–1996 for both pooled-gender and by-gender samples, although the trend is likely overstated by the survey redesign.[19]

Based on these comparisons of data and methods, we offer the following conclusions. First, estimates of educational differentials are quite consistent across data sources, sub-samples, and earnings concepts. Second, for most inequality outcomes, trends in full-time weekly earnings and overall hourly earnings are largely comparable within any given data source and are not particularly sensitive to the weighting scheme employed (bodies, weeks, or hours). Third, inferences regarding the residual distribution of earnings are far less consistent in sign, magnitude, and timing among data sources and are sensitive to the handling of outliers and selection of sub-samples. Although all data sources point to a growth of residual inequality starting in the 1970s, the relative magnitude, precise timing, and sample-specificity of this trend are elusive. These vagaries are unfortunate because shifts in the residual earnings distribution are less well understood than 'between group' inequality and, moreover, account, for the preponderance of recent inequality growth by most estimates. To make further progress in understanding these trends, researchers should carefully explore the robustness of their conclusions to choice of data source, sub-sample, and methodology.

2.3. Total compensation inequality versus wage inequality

A sharp increase in US wage inequality from the late 1970s to the mid-1990s is a well-documented and robust finding across a wide variety of datasets and studies. But wages do not represent the full economic returns to work. Non-wage employee benefits (fringe benefits), such as employer pension contributions and employer-provided health insurance, represent a significant share of total (pecuniary) compensation in the United States. Aggregate data from the National Income and Product Accounts indicates that supplements to wages and salaries as a percentage of total compensation increased rapidly from 7.5% in 1959 to 16.5% in 1979 to 18.9% in 1994, before declining slightly to 17.9% in 1996 (Economic Report of the President, 1998, Table B28, p. 312). Pierce (1997), using a somewhat broader measure of employee benefits, estimates that non-wage compensation represented 27.3% of total employer compensation costs in 1994. The non-pecuniary returns to work (working conditions) also vary substantially among jobs and individuals.

[19] Inequality measures make discreet upward jumps in 1994 in the ORG and 1993 in the March CPS, coincident with the redesign of the survey.

Table 4
Annualized changes in overall and residual inequality measures: 1960–1996 (10 × annualized changes)[a]

	Males						Females						Males & females					
	Overall			Residual			Overall			Residual			Overall			Residual		
	90–10	50–10	Var	90–10	50–10	Var	90–10	50–10	Var	90–10	50–10	Var	90–10	50–10	Var	90–10	50–10	Var
A. Full-time weekly earnings																		
1960s																		
March CPS	−0.03	−0.11	0.02	−0.01	−0.01	0.01	−0.03	−0.03	0.02	−0.02	−0.03	0.01	0.02	0.03	0.03	−0.01	−0.02	0.01
Census PUMS	0.10	0.03	0.03	0.03	0.01	0.02	0.05	−0.00	0.03	0.01	−0.03	0.02	0.07	−0.02	0.04	0.02	0.01	0.02
1970s																		
March CPS	0.10	0.11	0.03	0.11	0.08	0.03	0.03	−0.01	0.01	0.07	0.01	0.02	0.06	0.00	0.02	0.09	0.05	0.02
Census PUMS	0.10	0.11	0.05	0.09	0.05	0.04	−0.03	−0.11	−0.02	−0.03	−0.06	−0.02	0.07	0.02	0.02	0.05	0.01	0.01
May CPS	0.01	0.10	0.03	0.11	0.08	0.03	−0.12	−0.12	−0.02	0.01	−0.08	−0.00	0.02	0.05	0.01	0.07	0.01	0.02
1980s																		
March CPS	0.20	0.09	0.08	0.12	0.06	0.05	0.25	0.16	0.08	0.15	0.10	0.05	0.14	0.09	0.05	0.13	0.07	0.04
Census PUMS	0.17	0.06	0.03	0.07	0.02	0.00	0.16	0.10	0.04	0.09	0.05	0.01	0.08	0.02	0.01	0.08	0.03	0.00
CPS ORG	0.26	0.10	0.09	0.15	0.08	0.05	0.30	0.19	0.08	0.18	0.12	0.05	0.17	0.08	0.06	0.16	0.10	0.05
1990s																		
March CPS	0.11	−0.03	0.05	0.07	0.03	0.02	0.12	0.04	0.05	0.07	−0.01	0.02	0.09	−0.01	0.03	0.07	0.01	0.02
CPS ORG	0.05	0.00	0.05	0.06	0.02	0.03	0.12	0.04	0.06	0.07	0.02	0.03	0.07	0.03	0.04	0.06	0.02	0.03

B. All hourly earnings (weighted by hours worked)

1960s																		
March CPS	−0.01	−0.05	0.02	−0.01	−0.03	0.00	0.02	−0.01	0.00	−0.01	−0.02	−0.00	0.04	−0.04	0.02	−0.02	−0.04	−0.00
Census PUMS	0.04	−0.03	0.02	−0.01	−0.03	0.01	−0.01	−0.07	0.01	−0.01	−0.05	0.01	0.03	−0.03	0.02	−0.01	−0.03	0.01
1970s																		
March CPS	0.08	0.07	0.03	0.07	0.04	0.01	0.03	−0.03	0.01	0.05	−0.00	0.01	0.04	−0.01	0.01	0.05	0.02	0.01
Census PUMS	0.08	0.05	0.04	0.05	0.03	0.02	−0.11	−0.12	−0.06	−0.09	−0.10	−0.05	0.02	−0.04	−0.01	−0.00	−0.02	−0.01
May CPS	0.06	0.06	0.01	0.09	0.05	0.01	−0.11	−0.12	−0.03	−0.01	−0.08	−0.01	0.02	−0.04	−0.01	0.05	−0.01	0.00
1980s																		
March CPS	0.19	0.09	0.07	0.10	0.06	0.04	0.20	0.15	0.07	0.13	0.09	0.04	0.12	0.06	0.05	0.11	0.07	0.04
Census PUMS	0.14	0.05	0.03	0.07	0.02	−0.00	0.17	0.11	0.03	0.10	0.07	0.00	0.10	0.06	0.01	0.08	0.04	−0.00
CPS ORG	0.17	0.08	0.07	0.12	0.07	0.04	0.28	0.19	0.07	0.18	0.13	0.05	0.14	0.08	0.05	0.15	0.10	0.04
1990s																		
March CPS	0.12	−0.00	0.04	0.05	−0.01	0.03	0.16	0.01	0.07	0.08	0.02	0.05	0.08	0.03	0.04	0.06	−0.00	0.03
CPS ORG	0.07	−0.03	0.05	0.09	0.03	0.04	0.08	0.03	0.06	0.09	0.05	0.05	0.10	0.02	0.05	0.09	0.04	0.04

[a] Numbers above are 10 × annualized changes in earnings inequality metrics. See the notes to Table 3 for details on sample criteria and use of weights. Residuals are estimated from separate log earnings regressions in sample and years which include 9 education category dummies corresponding to years of school or highest degree completed (0, 1–4, 5–8, 9, 10, 11, some college, college graduate, post-college), a quartic in experience, and interactions between the experience quartic and dummies for less than high school, some college, and college or greater education. Pooled gender models also include a female dummy and interactions between female and the experience quartic and experience-education interaction terms.

The interpretation and welfare consequences of rising wage inequality clearly depends on whether it represents increased inequality in the overall economic returns to work as opposed to a change in the distribution of the composition of total compensation between wage and non-wage components. Thus a crucial research question is the extent to which changes in wage inequality are a good proxy for changes in the dispersion of the total economic returns to work. Research on changes in the distribution of the overall economic returns to work has been hampered by a lack of individual-level datasets with information on the incidence and value of non-wage benefits and by the difficulties involved in measuring and valuing non-pecuniary working conditions.

Pierce (1997) represents the most comprehensive study of the inequality of total hourly compensation (wage plus non-wage benefits) for the United States. Pierce examines reasonably representative national samples of jobs for 1986 and 1994 using the establishment survey micro data collected to produce the Employment Cost Index (a quarterly index of total employer compensation costs). This data provides information on hourly wages and on the incidence and value (employer cost) of a wide range of both legally required and voluntary benefits. Pierce finds that cross-sectional compensation inequality is greater than wage inequality. High wage jobs are more likely to have specific benefits (especially employer-provided health insurance, pensions, and paid leave) and a greater value of benefits. The differences in the incidence of voluntary benefits is especially large in the bottom-half of the wage (or total compensation) distribution. Pierce estimates a 90–10 log hourly compensation differential of 1.75 in 1994 as compared to a 90–10 log hourly wage differential of 1.568. Thus the cross-section data is suggestive of strong income effects in the demand for benefits with the benefit share increasing in total compensation. Furthermore Pierce's examination of data from 1986 to 1994 indicates a somewhat larger rise in compensation inequality than in wage inequality, especially in the bottom half of the compensation distribution.

Information on the incidence (but not on the valuation) of employer-provided health insurance and pension coverage is periodically available for nationally representative samples of employees from the Current Population Survey. These data indicate that changes in the incidence of employer-provided health insurance and pension coverage have exacerbated relative wage changes with a substantial decline in the relative likelihood of coverage for less-educated and low-wage workers from 1979 to the mid-1990s (e.g., Bloom and Freeman, 1992; Even and McPherson, 1994; Mishel et al., 1997a). For example, Farber and Levy (1998) document that the fraction of workers with health insurance from their own employer declined from 0.67 in 1979 to 0.50 in 1997 for high school dropouts as compared to a decline from 0.81 to 0.76 over the same period for college graduates.

Hamermesh (1999) provides a fascinating initial attempt to examine changes in the inequality of (non-pecuniary) workplace amenities. Hamermesh examines patterns of changes in inter-industry differentials in both wages and the total burden of occupational injuries from 1979 to 1995. He finds a widening of cross-industry inequality in the total burden of injuries with a relative drop in injuries in industries with rising relative earnings.

Hamermesh similarly finds in analysis of the timing of work from 1973 to 1991 that the incidence of work at unattractive hours (evenings and nights) has increased relatively for low-wage workers. Changes in the distribution of these workplace amenities also move in the direction of greater inequality in the total economic returns to work in the United States over the last two decades.

In summary, the limited available evidence strongly indicates that changes in the distribution of non-wage benefits and non-pecuniary workplace amenities tend to reinforce rather than offset observed increases in US wage inequality and wage differentials by education. This is an important area for future research, but a tentative conclusion is that recent changes in the wage distribution provide a reasonable proxy for changes in the distribution overall distribution of economic returns to work.

2.4. *Observable and unobservable components of changes in wage inequality*

Models of wage structure changes emphasizing shifts in the supply and demand for different labor inputs are likely to be easier to implement and interpret when applied to changes in relative wages among workers classified by observable skill categories. It is more difficult to separate out the contribution of changes in skill prices and quantities to changes in residual wage inequality. This raises the question of the extent to which changes in wage inequality reflect changes in the relative price and quantities of observed worker attributes as opposed to changes in residual inequality.

A common approach to assessing the quantitative contributions of observable and unobservable components of wage dispersion to changes in overall wage inequality is a standard variance decomposition. We start with a simple wage equation of the form

$$Y_{it} = X_{it}B_t + u_{it}, \tag{1}$$

where Y_{it} is the log wage of individual i in year t, X_{it} is a vector of observed individual characteristics (e.g., experience and education), B_t is the vector of estimated (OLS) returns to observable characteristics in t, and u_{it} is the log wage residual (which depends on the prices and quantities of unobserved skills, measurement error, and estimation error). The orthogonality of the predicted values ($X_{it}B_t$) and the residuals (u_{it}) in an OLS regression implies the variance of Y_{it} can be written as

$$\mathrm{Var}(Y_{it}) = \mathrm{Var}(X_{it}B_{it}) + \mathrm{Var}(u_{it}). \tag{2}$$

Thus the variance of log wages can be decomposed into two components: a component measuring the contribution of observable prices and quantities and the residual variance (a component measuring the effect of unobservables). These two components are typically referred to as between-group and within-group inequality. The change in variance of log wages between two periods can similarly be decomposed (by differencing Eq. (2)) into the change in the variance in the predicted values (change in between-group inequality) and the change in the residual variance (change in within-group inequality). This approach provides a clean and clear decomposition of wage inequality into observables and unob-

servables. The shortcoming of a reliance only on this approach is that the variance may not be the only inequality metric of interest especially given the sensitivity of the variance to changes in the tails of the distribution.

Table 5 presents such a between- and within-group decomposition of the growth of the variance of log weekly wages from 1963 to 1995 for our basic March CPS samples of full-time, full-year workers. Changes in the between-group variance component for men and for women reflect changes in relative returns to and the distribution of quantities of workers by education, experience, race, and region. The growth of residual inequality accounts for about a 60% of the increase in the variance of log weekly wages for both men and women over the full 1963–1995 period. This pattern reflects a somewhat more rapid proportional growth in between-group than residual inequality. In fact, for males the share of overall variance explained by the observables rises from the 32% in 1963 to 36% in 1995. The narrowing of the gender wage differential since 1979 reduces the between-group variance and implies a quite large contribution (75%) of residual inequality to the growth in overall wage inequality for men and women combined. The between group component plays a much larger role in the period of

Table 5
Between- and within-group components of changes in the variance of log weekly wages, full-time, full-year workers, March CPS 1963–1995[a]

	Changes in the variance components				
	Total change	Between-group change	Within-group change	% Explained	% Residual
A. Males					
1963–1995	0.159	0.067	0.092	42	58
1963–1979	0.047	0.014	0.033	33	67
1979–1995	0.112	0.053	0.059	47	53
B. Females					
1963–1995	0.131	0.048	0.083	37	63
1963–1979	0.022	−0.001	0.023	−5	105
1979–1995	0.109	0.049	0.060	45	55
C. Males and females					
1963–1995	0.111	0.028	0.083	25	75
1963–1979	0.037	0.010	0.027	27	73
1979–1995	0.074	0.018	0.056	24	76

[a] The between-group components (predicted values) and within-group components (residuals) of the variance of log weekly wages are based upon separate regressions by sex in each year of log weekly wages on 8 education dummies, a quartic in experience, 3 region dummies, black and other race dummies, and interaction between the experience quartic and 3 broad education category dummies. The regressions for males and females combined include the same covariates, plus a female dummy, and interactions of the female dummy with all other covariates.

rising educational differentials and accounts for 47% of the growth in male wage inequality from 1979 to 1995.[20]

Increases in between-group and within-group inequality are both important contributors to rising US wage inequality over the last several decades. A full explanation for changes in wage inequality needs to account not only for changes in returns to observed skill measure, but also for large changes in within-group inequality.

A further issue concerning the decomposition of changes in wage inequality into observable and unobservable components is the extent to which changes in between-group wage inequality reflects changes in the returns to observed skills as opposed to changes in the distribution of worker characteristics. The full-sample distribution accounting scheme developed by Juhn et al. (1993) is a useful approach that allows one to make such assessments for any measure of inequality (not just the variance). This approach begins with a simple wage equation such as (1) and conceptualizes the wage equation residual u_{it} as having two components: an individual's percentile in the wage distribution θ_{it} and the distribution function of the residuals $F_t(\ )$. By the definition of the cumulative distribution function, we can write the residual as

$$u_{it} = F_t^{-1}(\theta_{it} \mid X_{it}), \tag{3}$$

where $F_t^{-1}(\cdot \mid X_{it})$ is the inverse cumulative residual distribution for workers with characteristics X_{it} in year t.

The framework given by Eqs. (1) and (3) decomposes changes in inequality into three sources: (1) changes in the distribution of individual characteristics (changes in the distribution of the X's); (2) changes in the returns to observable skills (changes in the B's); and (3) changes in the distribution of residuals. By defining β as the average returns to observables over the whole period under study and $G(\cdot \mid X_{it})$ to be the average cumulative distribution, we can decompose the level of inequality into corresponding components using

$$Y_{it} = X_{it}\beta + X_{it}(B_t - \beta) + G^{-1}(\theta_{it} \mid X_{it}) + [F_t^{-1}(\theta_{it} \mid X_{it}) - G^{-1}(\theta_{it} \mid X_{it})]. \tag{4}$$

The first term captures the effect of changing distribution of worker characteristics; the second measures the effects of changing skill returns; and the third term accounts for changes in the distribution of the residuals. This framework allows one to reconstruct the (hypothetical) wage distribution that would attain with any subset of the components held fixed. One does not need to hold any of the components fixed at the average level for the

[20] The estimates of Juhn et al. (1993) similarly imply that an increase in residual wage inequality accounted for approximately 61% of the rise in the variance of log weekly wage for full-time, adult, white males in the March CPSs from 1964 to 1988. They also find a much larger contribution of the between-group component in the 1980s. DiNardo et al. (1996) find using data on hourly wages of all employees aged 16–65 from the CPS ORG samples that the majority (57%) of the increase in wage inequality from 1979 to 1988 is accounted for by rising between-group variance.

entire sample, one could simulate hypothetical wage distributions using any base period and replace β and $G(\cdot \mid X_{it})$ with the values for a reference period of interest.

If observable skill returns and the residual distribution are held fixed so that only observable quantities are allowed to vary, then wages would be determined by

$$Y_{it}^1 = X_{it}\beta + G^{-1}(\theta_{it} \mid X_{it}). \tag{5}$$

If observable skill returns and quantities are allowed to vary over time with only the residual distribution held fix, then wages are generated by

$$Y_{it}^2 = X_{it}B_t + G^{-1}(\theta_{it} \mid X_{it}). \tag{6}$$

The recommended approach of Juhn et al. (1993) is to calculate the distributions of Y_{it}^1, Y_{it}^2, and Y_{it} for each year studied and to attribute the change through time in the Y_{it}^1 distribution to changes in observable quantities. Any additional change in inequality in Y_{it}^2 beyond inequality changes in Y_{it}^1 is attributed to observable skill returns. Further change in actual overall inequality of Y_{it} beyond those found in Y_{it}^2 is attributed to residual inequality (changes in the distribution of residuals). The advantage of this approach over a standard variance decomposition is it allows one to look at how changes in each component affected the entire wage distribution and not just the variance. A disadvantage of moving away from the variance and examining other measures of inequality, such as quantile measures like the 90–10 log wage differential, is that these alternative measures typically do not uniquely decompose into between and within components. The actual allocations of changes in inequality to different components using the full sample accounting scheme are sensitive to the order in which one does the decomposition. The order chosen implicitly implies an assignment of interaction terms among the different components. Further ambiguities can arise since the specific results also depend on the base period chosen to hold components of the wage distribution fixed.[21]

Juhn et al. (1993) have implemented this approach for several quantile measures of wage dispersion using March CPS data on adult white males for 1964–1988. Table 6 summarizes their findings for the 90–10 log weekly wage differential. Increases in residual inequality account for 56% of the rise (0.208 of an increase of 0.373) of the 90–10 log weekly wage differential from 1964 to 1988. The contribution of residual inequality to the rise in the 90–10 differential is quite similar to findings from a standard variance decomposition. Table 6 also indicates that almost 80% of the contribution of observables to rising inequality for the whole 1964–1988 period result from increases in returns to observable skills (experience and education). In fact, the increase in returns to observed skills (mainly rising educational wage differentials) accounts for the majority (55%) of the increase in male wage inequality in the 1980s. Juhn et al. (1993) also report that increased returns to observed skills are more important for the increases in wage inequality in the upper half of

[21] For example, Goldin and Margo (1992) find substantial sensitivity of results to the choice of a base period in using this approach to decompose changes in US wage inequality from 1940 to 1950.

Table 6
Observable and unobservable components of changes in the 90–10 log wage differential, White males, March CPS, 1964–1988[a]

	Total change	Observed quantities	Observed skill returns	Unobservables
1964–1988	0.373	0.035	0.128	0.208
1964–1979	0.165	0.029	0.014	0.119
1979–1988	0.208	0.006	0.114	0.089

[a] Source: Juhn et al. (1993, Table 4).

the wage distribution than in the bottom half of the wage distribution as might be expected from the large increase in returns to college and advanced degrees in the 1980s.

2.5. Permanent and transitory components of earnings inequality

An increase in cross-sectional earnings inequality could reflect a rise in the permanent and/or the transitory component of earnings inequality. An explanation for the observed rise in cross-sectional inequality in the United States over the past several decades based on greater returns to skills (such as schooling and other persistent abilities) implies increased inequality in long-run (permanent) earnings. The substantial contribution of expanding educational wage differentials to growing earnings inequality is consistent with such a scenario. But the large increase in residual wage inequality could reflect increased returns to persistent (unobserved) worker attributes or a rise in transitory earnings variability. A sharp increase in the returns to (unobserved) skills is likely to have a much larger impact on long-run earnings inequality than an increase in transitory earnings instability. Explanations for increased wage inequality emphasizing the weakening of labor market institutions (e.g., unions, government wage regulation, internal labor markets) that increase the exposure of wages to market shocks may be consistent with increased year-to-year earnings turbulence. Understanding the contributions of changes in permanent and transitory components of earnings variation to increased cross-sectional earnings inequality is helpful for evaluating alternative hypotheses for wage structure changes and for determining the likely welfare consequences of rising inequality.

Following Baker and Solon (1998) and Moffitt and Gottschalk (1995), a rudimentary model of earnings dynamics allowing for time-varying earnings inequality is given by

$$y_{it} = p_t \alpha_i + \lambda_t \nu_{it}, \tag{7}$$

where y_{it} is the log earnings of individual i in year t, α_i is individual i's permanent earnings component (assumed to be time-invariant in this simple framework) with variance σ_α^2, ν_{it} is the transitory earnings component with variance σ_ν^2, α_i and ν_{it} are orthogonal to each

other, and p_t and λ_t are time-varying factor loadings on the permanent and transitory components of earnings. One interpretation of this framework is that α_i reflects persistent worker skills and p_t reflects the time-varying skill price (returns to skill). This model implies the variance of y_{it} can be written as

$$\text{Var}(y_{it}) = p_t^2\sigma_\alpha^2 + \lambda_t^2\sigma_\nu^2. \tag{8}$$

Eq. (8) shows that an increase in either factor-loading generates an increased cross-sectional earnings dispersion. The nature of the change in inequality depends on which factor loading changes. A persistent rise in p_t increases long-run earnings inequality (earnings dispersion across individuals measured over a long horizon such as a decade or lifetime) as the relative labor market advantage of high skill workers is enhanced. An increase in λ_t without an increase in p_t increases cross-section earnings inequality by raising year-to-year earnings volatility, but there is no increase in the dispersion of long-run earnings. An increase in p_t essentially maintains the rank order of individuals in the wage distribution, but spreads them out further in a persistent manner. An increase in λ_t leads to more changes in individuals' order in the earnings distribution, but the changes are quickly undone.

Measures of earnings mobility, the rate at which individuals shift positions in the earnings distribution (i.e., transition across quantiles of the earnings distribution), are closely related to the importance of permanent and transitory components in earnings variation. A large contribution of the permanent component implies that individuals' earnings are highly correlated over time (those with low relative earnings in one year are likely to have low relative earnings in other years) and thereby implies low rates of earnings mobility. Thus the extent to which changes in cross-sectional earnings inequality are driven by the permanent or transitory component has implications for changes in mobility rates. A rise in inequality caused solely by an increase in the permanent component will be associated with a decline in mobility rates. A rise in transitory component alone will increase mobility rates. Equal proportional increases in the permanent and transitory components will leave mobility rates unchanged even though earnings instability (the variation in year-to-year changes in log earnings for a typical individual) will be increased.

Since increases in the factor loading for either the permanent or the transitory component in Eq. (7) raises the cross-sectional variance of y_{it}, information on the time pattern of the variance of y_{it} from repeated cross-sections is not sufficient to identify whether p_t or λ_t has changed. Information on individual-level autocovariances of earnings is necessary to sort out changes in the permanent and transitory components of variance (Baker and Solon, 1998). Thus longitudinal data on individual earnings histories are required to assess the contributions of permanent and transitory components of earnings variation to levels and changes in earnings inequality.

A burgeoning literature has attempted to examine the contribution of permanent and transitory components of earnings variation to recent changes in US earnings inequality using data from several longitudinal datasets (the PSID, NLSY, and March–March

matched files from the CPS).[22] A consistent finding across studies and datasets is that large increases in both the permanent and transitory components of earnings variation have contributed to the rise in cross-section earnings inequality in the United States from the late 1970s to the early 1990s. The increase in the overall permanent component consists of both the sharp rise in returns to education and a large increase in the apparent returns to other persistent (unmeasured) worker attributes. The rise in cross-sectional residual inequality for males (controlling for experience and education) in the 1980s seems to consist of approximately equal increases in the permanent and transitory factors (Moffitt and Gottschalk, 1995).

Gottschalk and Moffitt's (1994) simple decomposition of the change in the variance of log earnings from the 1970s to the 1980s for male household heads in the PSID provides an illustrative set of results. Gottschalk and Moffitt subdivide their data into two 9-year periods, 1970–1978 and 1979–1987. After adjusting earnings for lifecycle earnings growth (controlling for an experience profile), they calculate for each individual the mean of his log earnings over the 9-year period (permanent earnings) and the deviation of his log earnings from the mean in each year (transitory earnings). The variance of permanent log earnings in each 9-year period is the variance of these 9-year means across individuals. They calculate the variance of transitory log earnings by computing the variance of the nine transitory components separately for each individual and then averaging them across individuals.[23]

Table 7 summarizes some of the key findings of Gottschalk and Moffitt (1994). The permanent and transitory variances both increased by about 40% from the 1970s to the 1980s. The similar proportional increases in transitory and permanent variances imply little change in earnings mobility. Roughly two-thirds of the increase in earnings variance (for both annual and weekly earnings) from the 1970s to the 1980s is accounted for by the permanent component, but the rise in earnings instability is still quantitatively significant. The changes in permanent and transitory variance are of similar magnitude when one looks within education groups (controlling for much of the increase in returns to education). The increase in earnings instability appears largest for less educated workers.

The implicit model of earnings dynamics used by Gottschalk and Moffitt (1994) is quite restrictive. For example, recent research on earnings dynamics provides evidence of: (1) persistent heterogeneity across individual not only in their level of earnings but also in their lifecycle growth rates; (2) the possibility of an important random-walk component to

[22] Gottschalk and Moffitt (1994), Haider (1997), and Moffitt and Gottschalk (1995) examine adult males using the PSID. Buchinsky and Hunt (1996) examine young workers using the NSLY. Gittleman and Joyce (1995, 1996) examine adult males and females using March–March matched files from the Annual Demographic Files of the CPS. Baker and Solon (1998) provide a sophisticated study of male earnings dynamics and changes in earnings inequality using a rich longitudinal data set of income tax records for Canada. See OECD (1997) for a summary of evidence on recent a changes in earnings mobility among other advanced nations. Studies of earnings mobility tend to focus on measures of annual earnings.

[23] This approach could be justified by an earnings dynamics model such as Eq. (1) if p_t and λ_t are fixed within each 9-year period but allowed to differ across the two 9-year periods.

Table 7
Variances of permanent and transitory log earnings, 1970–1987[a]

Sample definition	Permanent variance				Transitory variance			
	1970–1978	1979–1987	Change	Percent change	1970–1978	1979–1987	Change	Percent change
Log annual earnings								
All	0.201	0.284	0.083	41	0.104	0.148	0.044	42
Years of completed education								
Fewer than 12	0.175	0.272	0.097	55	0.106	0.208	0.102	96
12 or more	0.161	0.216	0.055	34	0.081	0.123	0.042	52
16 or more	0.184	0.200	0.016	9	0.065	0.093	0.028	43
Log weekly earnings								
All	0.171	0.230	0.059	35	0.075	0.101	0.026	35

[a] Source: Gottschalk and Moffitt (1994, Table 1).

earnings; and (3) serial correlation in transitory shocks to earnings (e.g., Abowd and Card, 1989; Baker, 1997). But more sophisticated empirical analyses that use more realistic (and complicated) models of earnings dynamics reach similar conclusions of substantial contributions of both permanent and transitory variances to the rise in cross-sectional earnings variance and little change in earnings mobility rates (e.g., Moffitt and Gottschalk, 1995; Haider, 1997).

A complete explanation for the recent rise in US wage inequality needs to account for both a growth in transitory earnings volatility and a large increase in the permanent variance component that appears associated with higher returns to education and other persistent worker attributes. The rise of earnings instability appears to be a bit of a puzzle for hypotheses only emphasizing rising skill prices associated with increased growth in the demand for skills relative to the supply of skills. A period of rapid skill-biased technological change associated with the spread of computer-based technologies and new organizational practices could both increase the relative demand for skill and (at least in a transition period) generate greater earnings instability since firms are likely to have much initial uncertainty concerning the abilities of individual workers' to perform new tasks and adapt to a new organizational environment. Rodrik (1997) has argued that increased globalization and international capital mobility can also increase earnings instability by making labor demand curves more elastic so that shocks to product market prices have a larger impact on wages. An important agenda for future work is to attempt to examine the extent to which patterns of changes in transitory earnings variability are related to changes in technology, organizational and personnel practices, exposure to international competition, changes in domestic product market competition, and changes in unionization and other labor market institutions.

2.6. Cohort versus time effects in inequality and the returns to education

The interpretation of recent increases in educational wage differentials and of within-group inequality (at least the persistent component of residual inequality) as largely reflecting increases in the returns to skills is facilitated by the (implicit) assumption that the distribution of unobserved ability is relatively similar across successive labor market cohorts.[24] An alternative possibility is that increased wage inequality may arise from increased dispersion of unobserved labor quality within recent entry cohorts, possibly from increasingly unequal school quality and diverging social conditions across neighborhoods. A decline in the unobserved ability of those with less education relative to those with more education in younger cohorts could potentially imply a rise in education returns

[24] Card and Lemieux (1996) provide an interesting formal assessment of the extent to which an increase in the returns to a single index of skill can account for the observed pattern of changes in wage differentials by education and age and in residual wage dispersion for the United States during the 1980s. They find that such a "single-index" model of skills provides a fairly accurate, but overly simplified, description of wage structure changes for white men and white women from 1979 to 1989.

reflecting an increase in ability bias.[25] In other words, changes in the wage structure could reflect changes in the average quality of different groups of workers rather than changes in the average wage for groups of workers of fixed quality.

Under the assumption that quality is relatively fixed within cohorts after school completion and labor market entry, these considerations have motivated investigations of the extent to which changes in inequality and educational differentials reflect changes within as opposed to between cohorts. Juhn et al. (1993) examine within-cohort changes in overall wage inequality (the 90–10 log weekly wage differential) for 6-year experience cohorts of white men. They find little within-cohort change in inequality in the 1960s, modest increases in the early 1970s, and large increases in the 1980s. The time pattern of average within-cohort inequality changes closely track average within-experience group changes. And Murphy and Welch (1993b) show that average within-cohort changes in the college wage premium similarly closely follow average within-experience group changes with a modest increase in the late 1960s, a decline in the 1970s, and substantial increases in the 1980s. Within-cohort changes (time differences) in inequality (or educational wage differentials) eliminate fixed cohort effects but could represent age or time effects or both. Although one cannot separately identify the levels of cohort, age, and time effects without very strong assumptions, a differences-in-differences approach of comparing within-cohort changes for different cohorts going through the same age ranges in different time periods can eliminate age and cohort effects and leave only changes in the time effect (the change in inequality growth over time). For example, a comparison of the change in inequality in the 1980s for the cohort aged 25–29 in 1980 to the change in inequality in the 1970s for the cohort aged 25–29 in 1970 provides an estimate of the difference in the time effect for the 1980s to the time effect for the 1970s.

Thus the findings of Juhn et al. (1993) shows an accelerating increase in inequality with time from the 1960s to the 1980s that cannot be explained by any combination of age and cohort effects. The sharp swings in within-cohort changes in educational wage differentials across decades (and even shorter periods in which changes in labor force composition are quite small) also strongly suggest that fluctuations through time in the college wage premium largely reflect changes in the relative price of educated labor and are not artifacts of changes in the composition of the college and high school populations.

A key role for changes in skill prices in movements in US educational wage differentials does not imply the absence of cohort or "vintage" effects in the returns to education. An exploratory analysis by Card and Lemieux (1999) reject the hypothesis that the return to education is the same for different cohorts in the US labor market. Their findings are

[25] A distinctive but related alternative hypothesis is that estimated changes in educational wage differentials reflect changes in the returns to unobserved ability rather than changes in "true" returns to education (e.g., Cawley et al. (1998). Changes in the returns to unobserved ability could lead to changes in ability bias even with unchanging distributions of unobserved ability within and between cohorts and education groups. This is a difficult issue requiring strong and controversial identification assumptions, but our reading is that the limited available evidence suggests substantial increases in the US college wage premium in the 1980s even after attempting to account for a rise in returns to unobserved ability (e.g., Chay and Lee, 1996).

suggestive of changing cohort effects in the college wage premium especially among recent US entry cohorts.

The much larger rise (within-experience group) rise in the college wage premium for younger than older workers in the 1980s could be attributed to either such changing cohort effects or from the larger impact of labor market shocks on younger than on older workers. Freeman's (1975) "active labor market" hypothesis postulates that changes in labor market conditions (changes in the supply and demand for skills) show up most sharply for new entrants because more senior incumbent workers are partially insulated from shocks by internal labor markets. This hypothesis suggests one should find lagged responses to shocks in older cohorts. It also implies that similar wage structure changes by skills should be apparent for new entrants and for displaced workers.

2.7. Longer-term historical changes in the US wage structure

Many explanations for recent wage structure changes emphasize factors, such as skill-biased new technologies and reduced barriers to international economic transactions, that are sometimes characterized as sharp breaks from the past. But rapid technological progress and reductions in communications and transportation costs have characterized advanced market economies for a long historical period stretching back at least to the industrial revolution. This raises the issue of how wage structure changes over the past several decades fit into longer-term historical patterns. Individual-level data on earnings and worker characteristics from the decennial Census of Population allow one to make reasonably consistent comparisons of wage structure changes (particularly for full-time, full-year workers) over the 1940 to 1990 period.[26] Nevertheless the 1940 Census PUMS is the first nationally-representative sample with information on both earnings or educational attainment. Thus the analysis of wage structure changes prior to 1940 is greatly constrained by data limitations and requires a focus on changes in wage differentials by occupation and/or industry (e.g., Douglas, 1930; Cullen, 1956; Chiswick, 1979; Williamson and Lindert, 1980; Goldin and Katz, 1995, 1998).

Table 8 uses data on log weekly wages of full-time, full-year, non-agricultural workers from the Census PUMSs to summarize the evolution of overall wage inequality (as measured by the 90–10 log wage differential) and the college wage premium (as measured by the regression-adjusted wage differential between those with exactly 16 years of schooling and those with exactly 12 years of schooling) from 1940 to 1990.[27] The existence of a large number of outlier observations with extremely low weekly earnings (especially for women in 1940) motivates our presentation of overall inequality measures based on two different approaches to trimming this bottom tail. The first approach deletes

[26] Recent studies using the Census data to examine wage structure changes over the full 1940–1990 period include Autor et al. (1998), Juhn (1994), Juhn et al. (1996), and Murphy and Welch (1993a).

[27] The Census collects information on annual earnings in the previous calendar year. Thus the data in Table 8 actually cover the 1939–1989 period. We focus on non-agricultural workers given the difficulties in measuring agricultural earnings especially in the early Census samples.

Table 8
US wage structure changes, 1940–1990 full-time, full-year non-agricultural workers, Census PUMSs[a]

	Males 90–10 differential		Females 90–10 differential		All college/ high school differential
	1% sample	MW sample	1% sample	MW sample	1% sample
1940	1.47	1.41	1.79	1.32	0.427
1950	1.00	1.00	1.10	1.06	0.303
1960	1.09	1.10	1.13	1.02	0.367
1970	1.18	1.16	1.18	1.02	0.409
1980	1.32	1.28	1.15	1.10	0.365
1990	1.48	1.52	1.30	1.33	0.501

[a] Note: All estimates are for log weekly wages of full-time, full-year workers not employed in agriculture. The 1% sample deletes the lowest 1% of workers sorted by log weekly wage. The MW sample deletes all workers earning less than 1/2 of the contemporaneous Federal minimum wage. The college/high school wage differential is the (adjusted) differential in log weekly wages of workers with exactly 16 years of schooling (or only a bachelor's degree in 1990) to those with exactly 12 years of schooling in regression of log weekly wages on 8 education dummies, a quartic in experience, 3 region dummies, a non-white dummy, a female dummy, and interactions of the female dummy with all other covariates except the education dummies.

the lowest 1% (and leads to findings that are quite similar to no deletions), and the second approach (following Juhn, 1994)) deletes all individuals who earned less than half the contemporaneous Federal minimum wage. This second approach could potentially be misleading given substantial changes in the coverage and relative generosity of the Federal minimum wage over the period of study (especially from 1940 to 1950).

The most striking feature of the data presented in Table 8 is the tremendous narrowing of wage inequality for both men and women in the 1940s.[28] Wage inequality for men then rises in each subsequent decade with an acceleration of the pace of widening inequality in the 1980s.[29] The entire compression of the wage structure in the 1940s is undone by 1990. The pattern for women is roughly similar. The US wage structure in the 1990s appears to be more unequal than at any point of time at least since 1940. The college wage premium also declines substantially in the 1940s, rises modestly in the 1950s and 1960s, narrows in the 1970s, and then sharply expands in the 1980s. Juhn (1994) shows that a wide variety of measures of educational and occupational wage differentials evolve similarly to the college wage premium from 1940 to 1990.

Overall wage inequality and educational wage differentials have expanded greatly since 1950 despite rapid educational advance and a large increase in the relative supply of more-

[28] Goldin and Margo (1992) refer to the 1940s as the period of the "Great Compression."

[29] Juhn (1994) reaches similar conclusions in an analysis of weekly earnings of full-time, white males from 1940 to 1990.

educated workers. Thus strong secular increases in the relative demand for skills is likely to be an important component of any explanation for US wage structure changes. The sharp contrast between the pattern of wage compression in the 1940s (a period of rapid expansion of unions, extremely tight labor markets for less-skilled workers associated with World War II, and government intervention in the economy) and of widening inequality in the 1980s (a period of eroding unions and sharp declines in blue collar employment in manufacturing) is suggestive of the possible importance of both institutional factors and changes in the relative demands for and supplies of different skill groups.

The available evidence on occupational wage differentials indicates a substantial decline in the earnings of white collar workers relative to blue collar workers from 1890 to 1939 (Goldin and Katz, 1995). This decline in the white collar wage premium occurs almost entirely in the decade surrounding World War I (especially from 1914 to 1919). The widening of occupational wage differentials from 1950 to 1990 has been large enough to offset the Great Compression of the 1940s, but it has not undone the compression that occurred around World War I. Thus the occupational wage structure has probably narrowed over the past century. The decades surrounding the two World Wars account for almost all the egalitarian movements in the wage structure in the 20th Century. The sources of these seemingly persistent effects of changes occurring during the period of the World Wars is an important question for an understanding of the long-run evolution of the US wage structure. One possibility is that wars enable the erosion of customary wage differentials (Phelps-Brown, 1977). The precise timing of the large declines in occupational/educational wage premiums in the 1910s and 1940s may reflect special factors related to the wars, but their persistence may reflect the role of market forces related to rapid expansions of the relative supply of more-educated workers associated with the high school movement after World War I and the growth of higher education after World War II.

3. Changes in other advanced OECD countries

Have wage differentials by skill and overall wage inequality increased in other advanced countries since the late 1970s to the same extent they have in the United States? A number of recent studies have attempted to assemble as comparable as possible data across advanced nations to answer this question.[30] Thus, in this section, we provide only a brief summary of the basic patterns of wage structure changes among advanced OECD nations over recent decades.

Table 9 classifies 12 countries by the way their educational and/or occupational wage differentials changed in the 1970s and the 1980s. During the 1970s, all the countries shared a common pattern of narrowing wage differentials by skill. Overall wage dispersion for males also narrowed in all of these countries with the exception of the United States. The

[30] See, for example, Berman et al. (1998), Davis (1992), Freeman and Katz (1994, 1995), Gottschalk and Smeeding (1997), Haskell and Slaughter (1998), and OECD (1993, 1996, 1997). The chapter by Layard and Nickell (this volume) examines cross-country differences in labor market institutions and labor market performance.

Table 9
Changes in educational/occupational skill differentials in selected countries[a]

Countries that experienced:	1970s	1980s
Large fall in differentials	Australia Canada France Germany Italy Japan Netherlands Sweden South Korea United Kingdom United States	Korea
Modest changes in differentials		
Modest fall in differentials		Netherlands
No noticeable change in differentials		France Germany Italy
Modest rise in differentials		Australia Canada Japan Sweden
A large rise in differentials		United Kingdom United States

[a] Source: Freeman and Katz (1994, 1995).

trend toward reduced educational wage differentials stopped or strongly reversed itself by the mid-1980s in all of these countries (except South Korea).

Furthermore patterns of changes in educational wage differentials and overall wage inequality are much more divergent in the 1980s and 1990s than in the 1970s. Table 10 measures changes in overall wage inequality for men from 1979 (or the earliest year available) to 1994 (or the latest year available) in terms of the 90–10 log wage differential. The United States and the United Kingdom experienced sharp increases in overall wage inequality, residual wage inequality, and, educational and occupational wage differentials of similar magnitude (Katz et al., 1995). The pattern of declining wage inequality apparent throughout the OECD (except the United States) in the 1970s ceased in the 1980s and 1990s in almost all nations (with Germany and Norway as possible exceptions). Canada, Australia, Japan, and Sweden had modest increases in wage inequality and educational/occupational differentials starting in the early 1980s.

Wage differentials and inequality narrowed through the mid-1980s in Italy and France with some hint of expanding in France in the late 1980s and with a large increase in

Table 10
Trends in wage inequality for males, selected OECD countries, 1979–1994[a]

Country	Log of ratio of wage of 90th percentile earner to 10th percentile earner				
	1979	1984	1989	1994	Change from earliest to latest year
Australia	1.01	1.01	1.03	1.08	0.07
Austria[b]	0.97		1.00		0.03
Canada[c]	1.24	1.39	1.38	1.33	0.09
Finland[d]	0.89	0.92	0.96	0.93	0.04
France	1.22	1.20	1.25	1.23	0.01
Germany[e]		0.87	0.83	0.81	−0.06
Italy	0.83	0.83	0.77	0.97	0.14
Japan	0.95	1.02	1.05	1.02	0.07
Netherlands[f]		0.92	0.96	0.95	0.03
New Zealand[g]		1.00	1.12	1.15	0.15
Norway[h]	0.72	0.72	0.77	0.68	−0.04
Sweden[i]	0.75	0.71	0.77	0.79	0.04
United Kingdom	0.90	1.02	1.12	1.17	0.27
United States	1.16	1.30	1.38	1.45	0.29

[a] Source: OECD (1996, Table 3.1, pp. 61–62). Notes: The samples generally consist of full-time workers, with the exceptions of Austria, Italy, and Japan. See OECD (1996, pp. 100–103) for details on the samples and earnings measures.

[b] Data for Austria in the 1979 column are for 1980.

[c] Data for Canada are for 1980, 1986, 1990, and 1994

[d] Data for Finland are for 1980, 1983, 1989, and 1994.

[e] Data for Germany are for 1983, 1989, and 1993.

[f] Data for the Netherlands are for 1985, 1989, and 1994.

[g] Data for New Zealand are for 1984, 1990, and 1994.

[h] Data for Norway are for 1980, 1983, 1987, and 1991.

[i] Data for Sweden are for 1980, 1984, 1989, and 1993.

inequality in Italy in the 1990s following the abolition of an automatic cost-of-living index favoring low-wage workers (the scala mobile) and the ending of synchronization of bargaining across industries. New Zealand also shows large increases in inequality in a period following substantial deregulation of product and labor markets (OECD, 1996).

These patterns are suggestive of an important role of differences and changes in labor market institutions and regulations in explaining the cross-country divergence of wage structure changes in the 1980s and 1990s. But differences in supply and demand factors may also play a role (e.g., greater decelerations in the rate of growth of relative skill supply growth in the United States and Great Britain from the 1970s to the 1980s). And the existence of either a decline in the relative wages of the less skilled, a sharp rise in the unemployment of the less skilled, or both in almost all OECD countries over the past two decades despite expanding relative supplies of highly educated workers is strongly sugges-

tive of a common shift in labor demand against the less skilled (Katz, 1994; Wood, 1994; Nickell and Bell, 1995). We next develop a framework to assess the roles of market forces and institutional factors in the evolution of national wage structures.

4. Conceptual framework: supply, demand, and institutions

This section develops a supply-demand-institutions (SDI) framework to assess the role of market forces (supply and demand shifts) and institutional factors in changes in the wage structure.

The specific approach taken borrows from the informal conceptual framework of Freeman and Katz (1994) and the more formal model of the determinants of between-group wage differentials of Bound and Johnson (1992).

The basic idea is that the actual wage of an individual can be decomposed into a latent "competitive" wage (or competitive total compensation level) and a deviation from the competitive compensation level for that individual. Actual wages may deviate from the competitive compensation level because of either institutional/non-competitive forces (unions, minimum wages, etc.) affecting wage setting or "measurement" problems arising from differences in non-wage compensation across jobs. The actual wage for individual i (w_i) can be defined as the product of the competitive wage for i (w_{ic}) and a relative rent for i (μ_i): $w_i = w_{ic}\mu_i$. If the non-wage employment attributes of all jobs were identical and there were no institutional or non-competitive factors causing wages to deviate from their competitive norm, then all the μ_i's would be equal to 1. But much evidence suggests that wages for given "quality" workers appear to systematically differ across industries and employers and by union status suggesting that deviations of μ_i from 1 are likely to be quantitatively important.[31] Deviations of wages from "full" competitive compensation whether arising from compensating differentials for non-wage attributes of employment or from non-competitive influences on wages are interpreted here as variation in relative rents.

This approach provides a useful framework for examining both changes in relative (log) wages among labor force groups and changes in residual (within-group) wage inequality. The aggregate work force is composed of K demographic groups (typically defined by age, education, and sex) indexed by k. The log wage for individual i in group k (Y_{ik}) can be expressed as the sum of the log competitive wage for i (Y_{ikc}) and the log relative rent for i (R_{ik}):

$$Y_{ik} = Y_{ikc} + R_{ik}, \tag{9}$$

where $Y_{ik} = \log(w_{ik})$, $Y_{ikc} = \log(w_{ikc})$, and $R_{ik} = \log(\mu_{ik})$. The mean log wage of group k

[31] Studies documenting and evaluating the evidence on inter-industry wage differentials include Slichter (1950), Krueger and Summers (1988), Katz and Summers (1989), Murphy and Topel (1990), and Gibbons and Katz (1992). Groshen (1991) examines US evidence on inter-employer wage differentials within detailed industries. Lewis (1986) carefully summarizes the US research on union/non-union wage differentials, and Card (1996) provides a thoughtful empirical analysis of differences in the "treatment" effect of unions on individual wages by skill group.

(the geometric mean of the wage rate of group-k workers) Y_k is conveniently equal to the sum of the competitive wage for group k (mean log competitive wage of group-k workers) and the average (log) rents for workers in group k

$$Y_k = Y_{kc} + R_k. \tag{10}$$

The competitive (log) relative wages (the Y_{kc}'s) are determined by the interaction of relative supplies and relative demands for the groups. To assist in the interpretation of the empirical literature, we concentrate on relative rents arising from three potentially measurable sources: (1) "true" industry wage differentials; (2) union wage effects; and (3) impacts of minimum wages or other forms of direct government intervention in wage setting. This focus leads us to also classify employment into J industries indexed by j.

The actual log wage of individual i of group k working in industry j is given by the sum of the competitive log wage for group k (Y_k); the mean industry wage differential (conditional on union status) for workers of group k employed in industry j (I_{jk}); a union status indicator ($U_{ik} = 1$ if i is unionized and 0 otherwise) times the associated mean union wage premium (λ_k) for group k; a minimum wage impact status indicator ($M_{ik} = 1$ if i's wage is affected by the minimum wage and 0 otherwise) and the associated mean minimum wage impact (δ_k) for affected workers in group k; and a (mean zero) individual error term (ε_{ijk}) reflecting measurement error and individual-level (within group) variation in ability and rents:

$$Y_{ijk} = \log(w_{ijk}) = Y_{kc} + I_{jk} + \lambda_k U_{ijk} + \delta_k M_{ik} + \varepsilon_{ijk}. \tag{11}$$

The industry wage differentials (I_{jk}'s) potentially reflect differential effects of unions on wage levels by industry and demographic group (differences in union bargaining power by industry, union threat effects, and union spillover effects), other sources of non-competitive wage variation across industries (efficiency wage and other rent sharing considerations), as well as equalizing differences for between-industry variation in working conditions and non-wage compensation. The mean minimum wage impact (δ_k) includes direct effects on for those earning the minimum wage as well as potential positive spillover effects above the minimum wage or possible negative crowding effects on wages in the uncovered sector.

The mean log wage for group-k workers can be written as

$$Y_k = Y_{kc} + \sum_j \{I_{jk}\phi_{jk} + \lambda_k U_k + \delta_k M_k\}, \tag{12}$$

where $\phi_{jk} = N_{jk}/N_k$ is the share of workers in group k that work in industry j; U_k is the fraction of group-k workers that are unionized; and M_k is the fraction of group-k workers that are affected by the minimum wage. We assume that log wages in each period are measured as deviations from the overall mean log wage. The change in the relative log wage of each group k is

$$\mathrm{d}Y_k = \mathrm{d}Y_{kc} + \sum_j (\mathrm{d}I_{jk}\phi_{jk} + I_{jk}\mathrm{d}\phi_{jk}) + \mathrm{d}\lambda_k U_k + \lambda_k \mathrm{d}U_k + \mathrm{d}\delta_k M_k + \delta_k \mathrm{d}M_k. \tag{13}$$

The relative wage of a particular group of workers can change either because market

forces lead its mean competitive wage to rise faster or slower than the overall average or because of changes in its relative rents. Eq. (13) indicates that changes in average relative rents for a group can arise from changes in the average level or incidence of industry wage premia, changes in the group's unionization rate or union wage premium, and changes in the impact of the minimum wage on that group.

Eqs. (9)–(12) analogously imply that changes in within group wage dispersion can arise from market forces affecting the distribution of competitive wages within a group (e.g., changes in the returns to unmeasured skills) or from institutional factors altering the within group distribution of rents (e.g., a change in the unionization rate for the group).

The SDI framework can be used to illuminate the strengths and weaknesses of the two primary empirical approaches to analyzing wage structure changes. The first approach assumes that changes in the wage structure largely reflect changes in competitive forces and uses a supply-demand model to explain actual relative wage and employment changes (e.g., Freeman, 1975; Katz and Murphy, 1992; Murphy and Welch, 1992). The basic idea is to see how far one can go with a pure competitive framework. The remaining "anomalies" can then be examined to determine the importance of institutional/non-competitive factors. The inherent difficulties in decomposing changes in within group wage dispersion into changes in prices and quantities means this approach is typically more straightforward to use in assessing the determinants of between group wage changes. The pure supply-and-demand approach can potentially be misleading to the extent exogenous institutional changes have a substantial effect on observed wages, especially if firms operate off their labor demand curves. Furthermore numerous difficult decisions arise concerning the appropriate level of aggregation of skill groups and strong assumptions are often required to separate out relative supply and demand shifts and to decompose measured relative demand shifts into interpretable factors such as the influences of skill-biased technological change, domestic product market demand shifts, and globalization factors (international trade and outsourcing). A more in-depth examination of the issues arising in the implementation of the supply-and-demand methodology and an assessment of the existing empirical literature using this approach is contained in Section 5.

The second approach more closely follows the framework illustrated in Eqs. (9)–(13) and tries to directly estimate the separate contributions of changes in institutional factors and competitive factors to observed changes in group relative wages and/or overall wage dispersion. The implementation of this approach to between-group wage differences typically uses relative wage change decomposition similar to Eq. (13) and involves three steps: (1) estimate the impact of changes in industry rents, union wage effects, and minimum wage influences on relative wages; (2) adjust actual wage changes for these institutional influences to uncover changes in relative competitive wages (the dY_{kc}'s); and (3) use an appropriate supply-demand model to examine the determinants of these changes in the structure of competitive wages. Bound and Johnson (1992) have developed an elegant framework to implement this methodology to account for between-group wage changes. DiNardo et al. (1996) have extended this approach to examine changes in overall,

between-group, and within-group wage dispersion; but their specific implementation limits the influence of supply and demand factors to only affecting between-group wage changes.

Two key issues arise in the implementation of the more direct SDI approach to sorting out institutional and competitive influences on the wage structure. The first is the issue of whether one can reliably estimate the direct influences of institutional/non-competitive factors on the wage structure and how these effects change over time. For example, this approach can generate misleading inferences of the influence of changes in industry rents to the extent estimates of industry wage differentials partially capture differences in unmeasured worker quality across industries (Gibbons and Katz, 1992; Murphy and Topel, 1990). And changes in minimum wages (real changes or changes relative to the median of the wage distribution) may not imply changes in the "bite" of the minimum wage if the underlying shadow competitive wages for low-wage workers are simultaneously changing. Furthermore estimates of union/non-union wage differential do not necessarily capture the full (general equilibrium) impact of unions on the wage structure, they provide estimates of differences in wages for given worker in union and non-union setting conditional on the current locus of unionization. Thus it is not clear how reliable existing estimates of union wage effects (or union effects on wage dispersion) are for doing counterfactuals of how the wage structure would differ if the locus of unionization were different. The attribution of wage structure movements to institutional changes may be problematic to the extent evolution of institutions reflects responses to market forces rather than exogenous events. A promising approach is to analyze wage structure changes associated with plausibly exogenous changes in institutions (e.g., the differential bite of changes in the Federal minimum wage across US states) or large discrete changes (e.g., deregulation or privatization of an industry or a major change laws affecting unions).[32]

The second related issue concerns the determination of employment when wages deviate from competitive levels. Even if one can adjust observed wage changes for institutional effects, observed employment changes are likely to depend (at least partially) on actual wages rather than on the latent competitive wages. Bound and Johnson (1992) attempt to conceptually escape this problem by assuming employment is set to equate marginal revenue products for each group to the group's underlying competitive wage. This assumption could be justified if deviations from competitive wages arise from union bargaining power and employers and unions negotiate over wages and employment to reach strongly "efficient bargains" (Farber, 1986). But much evidence suggests that even in union setting employment depends on actual negotiated wages rather than only on opportunity costs (e.g., Card, 1990) and this assumption is much less plausible for deviations from competitive wages caused by minimum wages.

Following Bound and Johnson (1991, 1992), we illustrate the operation of the SDI framework for assessing alternative explanations for between-group wage structure changes using a simple two group example. The work force is assumed to consist of

[32] See Fortin and Lemieux (1997) and Lee (1998) for recent attempts at using this approach

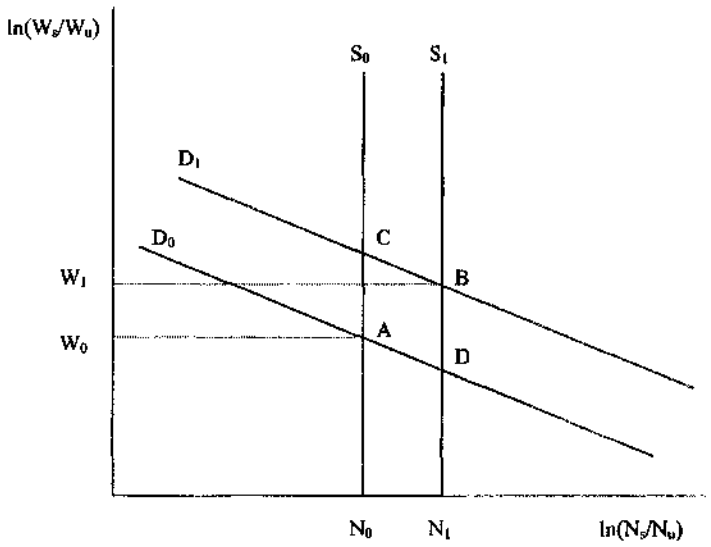

Fig. 6. SDI model.

two groups (skilled and unskilled workers). Data are available on actual log relative wages ($\log(w_s/w_u)$) and actual log relative employment ($\log(N_s/N_u)$) for two periods in which the relative wage and employment level of the more skilled group are both assumed to expand (perhaps representing wage structure and employment changes for college and non-college workers in the United States during the 1980s). Fig. 6 shows the economy moves from point A to point B. The question is to what extent does this observed change in relative wages and employment reflects the operation of competitive forces as opposed to institutional factors.

The pure supply and demand model assumes relative wages are determined by the intersection of the relative demand and supply curves in each period. Under the assumption of inelastic (predetermined) short-run relative supplies, the increase in the relative employment of skilled workers reflects a rightward shift in the relative supply of skilled workers in Fig. 6. If relative demand were stable, the relative wages of skilled workers would have declined. Thus an outward shift in the relative demand for skilled workers (from D_0 to D_1) must have been the driving force behind the rise in relative wage of skilled workers. This pattern leads analysts using a supply and demand model to focus on possible sources of demand shifts for the more-skilled (e.g., skill-biased technological change or product demand shifts across sectors with different skill intensities) and the variation in the rate of growth of relative skill supplies across time periods.

A possible institutional explanation for a rise in the skill differential is a decline in the relative rents of unskilled workers. In this case the rise in the relative wage and employment of the skilled from A to B in Fig. 6 could arise even with no shift in the relative

demand curve. For example, the relative demand curve could be stable at D_1, but unskilled workers initially received large rents from unions with firms setting employment at the competitive level. In this case, the economy initially operates off the labor demand curve at point A rather than C. The increase in the relative supply of the skilled would have reduced wages to point D, but the complete erosion of rents results in the increased skill premium at point B.

Of course a mixture of both a decline in relative rents and some shift in relative demand favoring the skilled could also be consistent with the observed change in relative wages and employment. Furthermore, the "naive" supply and demand analysis would correctly estimate the effects of demand shifts even in the presence of rents as long as wages are set equal to marginal products. When employment lies on the labor demand curve, wage changes arising from changes in rents affect unemployment (or non-employment rates). Thus knowledge of the slope of the relative demand and information on observed changes in relative wages and quantities would allow one to uncover relative demand shifts, but this approach could attribute wage changes to relative supply shifts that might reflect changes in relative rents. Information on changes in population shares or labor force shares by skill group potentially can be used to supplement relative employment information to sort out the effects of changes in relative skill supplies from changes in relative rents (e.g., Nickell and Bell, 1995; Jackman et al., 1997).

5. Supply and demand factors

This section develops the pure supply and demand approach to analyzing wage structure changes. We begin with a generic supply and demand framework to analyze between-group relative wage changes. We show how this framework can be used to assess whether observed changes in relative wages and relative employment are consistent with stable relative factor demands. We then examine key modeling issues concerning the specific approach to aggregating heterogeneous demographic groups into distinct labor inputs (skill groups) and assumptions concerning market clearing and the exogeneity of relative factor supplies. The framework is used to examine recent US wage structure changes. The importance of between versus within industry demand shifts and the roles of variation in the rate of growth of relative skill supplies, skill-biased technological changes, and globalization factors in changes in wage differentials by education are assessed.

5.1. A simple supply and demand framework

We begin by examining between-group relative wage changes using a simple supply and demand framework from Katz and Murphy (1992) in which different demographic groups (identified by sex, education, and age/experience) are treated as distinct labor inputs. The relative wages of demographic groups can be thought of as being generated by the interaction of the relative supplies of the groups and an aggregate production function with its

associated factor demand schedules. The determinants of relative factor supplies are not specified in the initial framework. The key requirement for this approach to be plausible is that observed factor prices and quantities must be "on the demand curve."

The basic framework posits an aggregate production function consisting of K types of labor inputs. We assume the associated factor demands can be written as

$$N_t = D(W_t, Z_t), \tag{14}$$

where $N_t = K \times 1$ vector of labor inputs employed in the market in year t, $W_t = K \times 1$ vector of market wages for these inputs in year t, $Z_t = m \times 1$ vector of demand shift variables in year t.

The demand shifters, Z_t, capture the effects of technology, product demand shifts, and other non-labor inputs on demands for labor inputs. Since we are concerned with explaining *relative* wage changes as a function of *relative* supply and *relative* demand shifts, we abstract from changes in absolute wages arising from factor-neutral technological change and from neutral demand shifts associated with changes in the scale of the economy. In practice W_t is a vector of relative wages where actual wages have been deflated by a fixed-weighted wage index capturing aggregate wage changes, and N_t is a vector of relative supplies measured as a share of total labor input in the economy in each year measured in *efficiency units*. Actual hours worked for each group are translated into efficiency units by multiplying by the average relative wage for group in some base period.[33]

Under the assumption that the aggregate production function is concave, the $(K \times K)$ matrix of cross-price effects on factor demands, D_w, is negative semidefinite. Eq. (14) can be written in terms of differentials as

$$\mathrm{d}N_t = D_w \mathrm{d}W_t + D_z \mathrm{d}Z_t. \tag{15}$$

Thus relative wage changes depend on changes in net relative supplies (relative supplies net of relative demand shifts)

$$\mathrm{d}W_t = [D_w]^{-1}(\mathrm{d}N_t - D_z \mathrm{d}Z_t). \tag{16}$$

The impact of changes in net relative supplies on relative wages depend on the degree of substitutability and complementarity among different labor inputs in the aggregate production function.

The negative semidefiniteness of D_w implies from Eq. (15) that

$$\mathrm{d}W_t'(\mathrm{d}N_t - D_z \mathrm{d}Z_t) = \mathrm{d}W_t' D_w \mathrm{d}W_t \leq 0. \tag{17}$$

Changes in factor quantities (net of demand shifts) and changes in wages must negatively covary if observed wages and quantities lie on the factor demand curves. If factor demand is stable (Z_t fixed), Eq. (17) implies $\mathrm{d}W_t' \mathrm{d}N_t \leq 0$. Actual changes in relative wages and relative quantities must negatively when factor demands are unchanging. In the case of

[33] Katz and Murphy (1992) and Murphy and Welch (1992) provide more detailed discussions of alternative approaches to measuring relative wages, relative factor supplies, and defining efficiency units.

two inputs, the intuitive basic implication of stable relative factor demand is that an increase in the relative supply of a group must lead to a reduction in the relative wage of that group. Furthermore data on relative factor quantities and wages alone can be used to assess whether observed wage structure changes over any period are consistent with a stable factor demand structure.

This approach can be illustrated using data on recent US relative wage and supply changes. Much early work examining US wage structure changes in the 1970s emphasized the role of "exogenous" relative supply shifts from changing demographics and school completion rates as the driving force behind relative wage changes (e.g., Freeman, 1979; Welch, 1979). This might appear to be a reasonable first approach for this period of the labor market entry of the US baby boom cohorts in which rapid expansions of the relative supply of more-educated and younger workers coincided with narrowing educational wage differentials and expanding experience differentials. But an examination of data since the late 1970s or over longer time periods clearly rejects the assumption of stable factor demands and implies an important role of demand shifts especially secularly rising relative demand for more-educated workers (e.g., Bound and Johnson, 1992; Katz and Murphy, 1992; Murphy and Welch, 1992; Johnson, 1997; Autor et al., 1998).

Data on relative supply changes for the United States by sex, education, and experience groups for 1963–1987 and several sub-periods from the March CPS are illustrated in Table 11. These relative supply changes can be compared to the relative wage changes for the same time periods shown in Table 2. Since the relative supplies and wages of more educated workers and females increased over this 25-year period, it is clear that relative demand shifts are necessary to explain the observed data. Katz and Murphy (1992) divide the labor force into 64 groups (defined by sex, education, and experience) and use estimates of the time series (N_t, W_t) covering the 1963–1987 period to assess the stable factor demand hypothesis between any given years t and year τ by evaluating whether

$$(W_t - W_\tau)'(N_t - N_\tau) \leq 0. \tag{18}$$

Time periods for which the inequality in (18) is satisfied (i.e., the inner product of changes in wages and changes in factor supplies is non-positive) have the potential to be explained solely by supply shifts. When this inequality is not satisfied, no story relying entirely on supply shifts is consistent with the data.[34] This inequality clearly fails for the entire 1963–1987 period as illustrated by the plot in Fig. 7.[35] Demand shifts favoring more-educated workers and women are necessary within this framework to explain the pattern of relative wage and quantity changes from 1963 to 1987. Expanding relative wages of more-skilled workers in the face of increased relative supplies of more-educated workers are also apparent in many other OECD nations in the 1980s and 1990s (OECD, 1993, 1996; Gottschalk and Smeeding, 1997).

[34] Murphy and Welch (1992) present a formal statistical framework for testing the stable factor demand hypothesis embodied in Eq. (18) and implement this framework on US data for men for 1963–1989.

[35] But Katz and Murphy (1992) and Murphy and Welch (1992) find that inequality (18) is satisfied for the 1970s.

Table 11
US relative supply changes, 1963–1987[a]

Group	Change in log share of aggregate labor input (multiplied by 100)			
	1963–1971	1971–1979	1979–1987	1963–1987
Gender				
Men	−2.9	−4.9	−4.2	−12.0
Women	11.2	15.7	11.2	38.2
Education (years of schooling)				
8–11	−35.2	−48.6	−41.9	−125.7
12	7.6	−4.8	−4.8	−2.0
13–15	20.3	23.3	6.7	50.3
16+	17.8	24.1	15.6	57.5
Experience (men)				
1–5 years	30.3	16.3	−27.9	18.6
6–10 years	14.2	19.5	−10.4	23.4
11–15 years	−4.3	6.9	17.5	20.1
16–20 years	−17.8	−6.6	22.7	−1.7
21–25 years	−15.5	−16.9	0.0	−32.3
26–35 years	−5.5	−23.8	−17.4	−46.7
Education 12				
Experience 1–5	16.2	18.7	−40.9	−6.0
Experience 26–35	4.0	−26.9	−10.9	−33.8
Education 16+				
Experience 1–5	52.7	17.1	−12.7	57.1
Experience 26–35	19.8	18.9	−5.8	32.9

[a] Source: Katz and Murphy (1992, Table 2). Notes: The numbers in the table represent log changes in each group's share of total labor supply measured in efficiency unites (annual hours times the average relative wage of the group for the 1963–1987 period) using data from the March Current Population Surveys for 1964–1988.

Relative demand shifts favoring more-skilled workers are also essential to understanding longer-run changes in the US wage structure. Table 12 displays the evolution of the educational composition of aggregate US labor input (for those aged 18–65 years) measured in full-time equivalents (total hours worked) and of the log college/high school wage differential from 1940 to 1996.[36] The educational attainment of the work force increased rapidly over this 56-year period with a more than fourfold increase in the share of hours worked by those with at least some college. Despite the large increase in the relative supply of the more educated, the college/high school wage differential has

[36] The large increases in the educational attainment of the US work-force since 1940 may overstate increases in the relative supply of "more-skilled" workers to the extent that the "unobserved" quality of more-educated workers declines with some "re-labeling" of "lower productivity" workers into higher education categories. Juhn et al. (1996) examine this issue using Census PUMS data from 1940 to 1990 and find that conclusions concerning changes in relative supply and implied relative demand shifts are not much affected by adjustments for such re-labeling through controls for cohort-specific college share or mean years of education.

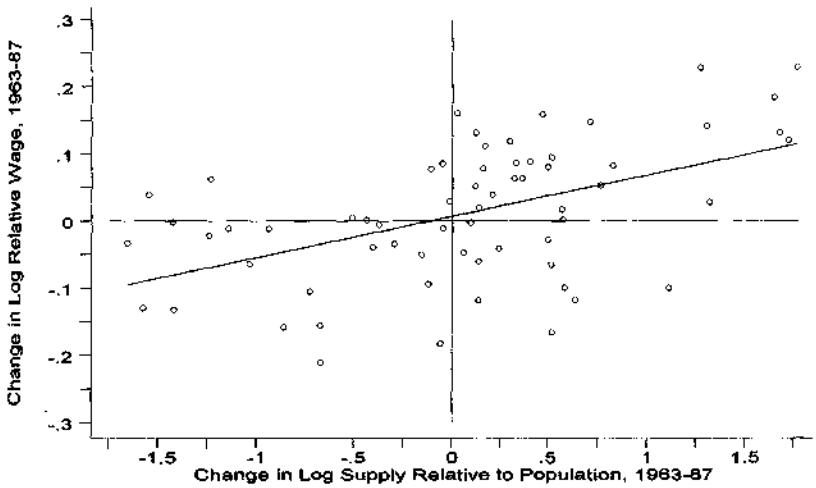

Fig. 7. Price and quantity changes for 64 groups, 1963–1987.

grown substantially since 1950 suggesting sharp secular growth in the relative demand for the more educated that started well before the rise in wage inequality of the 1980s.[37] But fluctuations in the rate of growth of the relative supply of more-educated workers also appear to have played an important role in the time pattern of changes in educational wage differentials. Tables 11 and 12 illustrate that an increase in the rate of growth in the supply of college workers in the 1970s was associated with a decline in the college wage premium and a decrease in the rate of growth of the supply of college workers in the 1980s was associated with a sharp rise in the college wage premium. A rather smooth trend increase in the relative demand for more-educated workers combined with observed fluctuations in the rate of growth of the relative supply has the potential to explain much of the evolution of US educational wage differentials at least over the past few decades.

The consistency of alternative hypotheses (alternative choices of demand shifters Z_t) concerning the evolution of relative demand with the observed pattern of changes in relative wages and supplies from τ to t can be assessed using a discrete version of Eq. (17)

$$(W_t - W_\tau)'[(N_t - N_\tau) - (D(W_\tau, Z_t) - D(W_\tau, Z_\tau))] \leq 0, \tag{19}$$

which involves evaluating the value of the inner product of the change in wages from year τ to year t with the changes in net supplies (equal to the actual change in relative factor supplies less the change in relative demands that would have happened at fixed factor

[37] Early papers by Griliches (1970) and Welch (1970) inferred substantial relative demand shifts for the more-educated in the 1950s and 1960s to explain the failure of the college wage premium to decline in the face of the rising relative supply of college workers.

Table 12
Educational composition of employment and the college + /high school wage premium, 1940–1996[a]

	Full-time equivalent employment shares by education level (%)				
	High school dropouts	High school graduates	Some college	College graduates	Log college +/ high school wage
1940 Census	67.9	19.2	6.5	6.4	498
1950 Census	58.6	24.4	9.2	7.8	313
1960 Census	49.5	27.7	12.2	10.6	396
1970 Census	35.9	34.7	15.6	13.8	465
1980 Census	20.7	36.1	22.8	20.4	391
1980 CPS ORG	19.1	38.0	22.0	20.9	356
1990 CPS ORG	12.7	36.2	25.1	26.1	508
1990 Census	11.4	33.0	30.2	25.4	549
Feb. 90 CPS	11.5	36.8	25.2	26.5	533
1996 CPS ORG	9.4	33.4	28.9	28.3	557

[a] Source: Autor et al. (1998, Table 1).

prices). A particular hypothesis of interest is whether that data are consistent with a stable trend rate of demand change for each labor force group with fluctuations in relative wages about trend driven by detrended relative supply changes. Such trend demand shifts might reflect a rather steady pace of non-neutral technological change or steady shifts in the industrial composition of employment. Katz and Murphy (1992) and Murphy and Welch (1992) find for US data that allowing for trend demand shifts virtually eliminates inconsistencies with otherwise stable demand for the overall period from the early 1960s to the late 1980s, but Katz and Murphy conclude that some acceleration of demand shifts favoring the more-educated and women in the 1980s is required to explain difference among sub-periods in the pattern of relative wage and employment changes.

Analyses of US changes in relative wages and factor supplies over recent decades using a simple supply and demand framework indicate a key role for strong secular shifts in the relative demand favoring the more skilled and decade-to-decade fluctuations in the pace of relative supply changes. An assessment of the quantitative importance for explaining relative wage movements of relative supply and demand shifts and of the underlying sources of the demand shifts requires adding more structure to the framework.

5.2. Some issues in supply and demand analysis

The assessment of whether economy-wide changes in relative wages and quantities employed are consistent with stable factor demand requires that aggregate factor demand equations (as in Eq. (14)) satisfy the usual properties of factor demands and that actual wages and employment levels lie on these factor demand equations.[38] No assumptions about the determinants of relative factor supplies are necessary.

Further progress on the contribution of different supply and demand factors to wage structure changes requires additional assumptions about the determinants of factor supplies and the functional form of the factor demand equations. Two key assumptions typically made are that of full-employment (relative wages adjust so that relative supplies equal relative demands) and exogenous (or at least pre-determined) relative supplies. Relative supplies are treated as pre-determined by past educational investment decisions and demographic changes arising from earlier fertility and immigration decisions. Current labor force participation decisions are assumed to be unaffected by current market conditions. Thus the basic model is one of a vertical (inelastic) short-run relative labor supply curve as in Fig. 6. Relative quantities employed are determined by pre-determined relative supplies, while both relative demand and supply factor affect relative wages.

The full employment/market clearing assumption may be reasonable for the United States, but it is clearly is problematic for examining European economies over the past two decades. Jackman et al. (1997) have extended the basic model to allow for bargaining factors and unemployment under the assumption that relative supply shifts can be measured by exogenous changes in relative labor force sizes by skill group. The well-documented decline in the relative employment/population ratios (through both rising relative unemployment rates and declining relative labor force participation rates) of groups with declining wages in the United States since the 1970s (e.g., Murphy and Topel, 1997; Murphy and Welch, 1997) further suggests the assumption of exogenous inelastic relative labor supply curves may also be problematic the United States. Relative population shares of different groups can potentially be used to instrument for relative employment shares to allow for an elastic short run supply curves if relative population shares by sex-education-age groups can plausibly be viewed as pre-determined.

Two other key decisions required to implement a supply and demand analysis are an assumed functional form of the factor demand schedules and a choice concerning how to disaggregate labor input into different skill groups. These decisions involve (explicit or implicit) assumptions about the nature of the aggregate production function.

Many alternative approaches to the aggregation of heterogeneous labor force groups into "appropriate" skill groups have been used in recent research on wage structure changes.[39] One would like to aggregate workers into groups such that workers are much closer substitutes in production within the groups than between the groups. The implicit assumption is that hours of work by different workers are perfect substitutes within a skill group. But the hours of different workers can easily be given different weights in adding up the total supply within a group such as through the approach of measuring labor supplies in efficiency units with each worker's hours weighted by the average wage in a base period of that worker's more detailed sub-group.

A fruitful first-cut approach that is easy to implement is to break up the work force into

[38] Thus such assessments may be inaccurate if relative wage changes are driven by institutional factors that force firms off their labor demand curves.

[39] Hamermesh (1993) provides a detailed and thoughtful discussion of the issues arising in the choice of an aggregation scheme in empirical work on labor demand.

two groups along the wage structure dimension of particular interest: high-education and low-education to examine educational wage differentials, "young" and "old" to study experience differentials, and men and women to examine gender differentials. The groups can typically be chosen so that the assumption of much greater substitutability within than between groups is plausible and estimates using such an approach are easy to interpret. The disadvantage is one loses much information about the subtleties of wage structure changes from this extreme approach to aggregation. Examples of this approach include the analyses of relative wage changes for two education groups, skilled (college or more) and unskilled (less than college) by Autor et al. (1998), Baldwin and Cain (1997) and Krussell et al. (1997). Much research has also analyzed wage structure and relative demand changes for two broad occupation groups such as production and non-production workers (e.g., Berman et al., 1994, 1998). Such a broad occupational breakdown is often all that is available for many datasets derived from establishment-based surveys such as the US Annual Survey of Manufactures or cross-country data for manufacturing industries from the UN General Industrial Statistics Database. The assumption of pre-determined relative supplies is clearly much less plausible for an occupational grouping than for education or age groupings. But Berman et al. (1994) and Machin and Van Reenen (1998) find that a non-production/production worker approach does a reasonable job of matching a high/low education group breakdown in manufacturing for most advanced industrial nations.

A hybrid of the two-group approach is to examine the relative wage of two "pure" skill classes (college graduates and high school graduates) and to relate this relative wage to changes in the relative supply and demands for "equivalents" of these pure skill classes (college and high school equivalents). The aggregation of multiple skill groups into two pure skill classes follows the "linear synthesis" approach developed by Welch (1969) by assuming each skill group is a linear combination of the two pure skill classes with the weights usually based on the extent to which wages of each group tracks those of the pure skill groups (e.g., Katz and Murphy, 1992).

The alternative approach is to specify labor input as consisting of a large number of possible inputs typically defined by sex, education, age/experience groups or with even further differentiation by race and foreign born status. The advantage of this approach is the ability to gain much more information about the nature of wage structure changes (e.g., differences in changes in educational wage differentials for older and younger workers, etc.). But strong assumptions about functional forms and substitution possibilities between groups must be imposed to make this approach feasible. Restrictions on substitution possibilities reduce the number of parameters to be estimated in the factor demand system to a practical number. A breakdown of the work force into K groups implies the matrix of cross-price elasticities among the groups (D_w in Eq. (15)) as well as the related substitution matrix ($[D_w]^{-1}$) both contain $K \times K$ elements implying an enormous number of separate parameters for large K even after imposing symmetry if one does not make further restrictions. The estimation of this many separate parameters for large K is unlikely to be feasible and will more than exhaust the available degrees of freedom when the number

of groups is large relative to the time periods or the cross-section units (different regions) being used as the source of identifying variation. For example, Bound and Johnson (1992) examine 32 demographic groups using data from 3 years and Murphy and Welch (1992) examine 188 groups over 27 years.

The first method to addressing this problem is to assume a particular functional form for the production function to limit the number of substitution parameters. Bound and Johnson (1991, 1992) assume a constant elasticity of substitution (CES) production function with each of 32 demographic groups as the inputs and thereby estimate a single intrafactor substitution parameter.[40] The key assumption underlying this approach is that the degree of substitutability in production of between any pair of groups is the same. Thus the degree of substitutability between young male high school graduates and high school dropouts is assumed to be equivalent to the degree of substitutability between young male high school dropout and experienced female college graduates. This assumption seems implausible given the similar occupational and industrial distributions of young male high school graduates and dropouts and the quite dissimilar occupational and industrial distributions of young male dropouts and experienced female college graduates (Murphy and Welch, 1997). But Bound and Johnson (1992) show a major advantage of the CES approach is that it can be applied at the sectoral level and provides an interpretable structural framework to analyze between- and within-industry demand shifts for multiple skill groups.

A second method is to aggregate the number of groups to a smaller feasible number to allow more general patterns of substitution among the groups (such as the three group approach of Jaeger (1995)). The third method is to assume that wages for individual workers depend on their quantities of a smaller number, $k < K$, of (latent) basic skills. The endowments of each of the k underlying skills for K groups vary at a point of time but are assumed to be stable over time. Murphy and Welch (1992) show how this approach greatly reduces the number of parameters to be estimated in the factor demand structure for small k and still allows a rich pattern of substitution possibilities among the K groups.[41]

5.3. Supply and demand analysis of changes in educational wage differentials

Many studies (at least since Freeman (1975)) have used simple supply and demand frameworks to analyze changes in educational wage differentials in the United States and other countries. A common approach is to break the work force into two broad educational groups.[42] We illustrate this approach by considering a CES production function for aggregate output Q with two factors, college equivalents (c) and high school equivalents (h):

[40] Card (1997) similarly uses a CES production function with ten skill deciles as the distinct inputs.

[41] Teulings (1997) develops an alternative approach to aggregation allowing for an infinite number of skill classes but adding structure based on an assumption of the comparative advantage of more skilled workers in more complex jobs.

[42] Empirical analyses of more general supply-and-demand frameworks to assess a range of wage structure changes (e.g., education, experience, and gender differentials) include Katz and Murphy (1992) and Murphy and Welch (1992).

$$Q_t = [\alpha_t(a_t N_{ct})^{\rho} + (1 - \alpha_t)(b_t N_{ht})^{\rho}]^{1/\rho}, \tag{20}$$

where N_{ct} and N_{ht} are the quantities employed of college equivalents (skilled labor) and high-school equivalents (unskilled labor) in period t, a_t and b_t represent skilled and unskilled labor augmenting technological change, α_t is a time-varying technology parameter that can be interpreted as indexing the share of work activities allocated to skilled labor, and ρ is a time invariant production parameter.

Skill-neutral technological improvements raise a_t and b_t by the same proportion. Skill-biased technological changes involve increases in a_t/b_t or α_t. Following Johnson and Stafford (1998), one can interpret increases in a_t/b_t as *intensive* skill-biased technological change in which skilled workers get relatively better at their existing jobs more rapidly than do unskilled workers. Increases in α_t can be viewed as *extensive* skill biased technological change or "upskilling" that shifts work tasks from unskilled to skilled workers.[43] The aggregate elasticity of substitution between college and high-school equivalents is given by $\sigma = 1/(1 - \rho)$.

Although the single-sector, aggregate production function directly including only labor inputs given in Eq. (20) is a well-defined analytical construct, one must be clear about what it means. Such an aggregate production function does not necessarily have any simple interpretation in terms of the production functions of individual firms or even industry-level production functions. The aggregate elasticity of substitution σ reflects not only technical substitution possibilities in firm-level production functions but also outsourcing possibilities and substitution possibilities across goods and services in consumption. Changes in the "technology" indicators a_t/b_t and α_t represent not only true technological changes at the firm level but also the non-neutral effects on skill groups of changes the relative prices or quantities of non-labor inputs (capital, energy) and shifts in product demand among industries with different skill intensities.

Under the assumption that college and high-school equivalents are paid their marginal products, we can use Eq. (20) to solve for the ratio of marginal products of the two labor types yielding a relationship between relative wages in year t, w_{ct}/w_{ht}, and relative supplies in year t, N_{ct}/N_{ht} given by

$$\log(w_{ct}/w_{ht}) = \log(\alpha_t/[1 - \alpha_t]) + \rho\log(a_t/b_t) - (1/\sigma)\log(N_{ct}/N_{ht}), \tag{21}$$

which can be rewritten as

$$\log(w_{ct}/w_{ht}) = (1/\sigma)[D_t - \log(N_{ct}/N_{ht})], \tag{22}$$

where D_t indexes relative demand shifts favoring college equivalents and is measured in log quantity units. The impact of changes in relative skill supplies on relative wages depends inversely on the magnitude of aggregate elasticity of substitution between the

[43] Goldin and Katz (1998) model and document this process of upskilling from less-skilled to more-skilled production workers and from production to non-production workers in the US manufacturing sector with the spread of electricity and adoption of continuous process and batch production methods from 1890 to 1929.

two skill groups. The greater is σ, the smaller the impact of shifts in relative supplies on relative wages and the greater must be fluctuations in demand shifts (D_t) to explain any given time series of relative wages for a given time series of relative quantities. Changes in D_t can arise from (disembodied) skill-biased technological change, non-neutral changes in the relative prices or quantities of non-labor inputs such as computer services, increased outsourcing possibilities that disproportionately affect the two skill groups, and shifts in product demand either from domestic or international sources.[44]

Two approaches can be taken using this framework to assess alternative stories for relative wage changes by skill group consistent with the observed pattern of changes in relative wages and quantities employed. The first is to directly estimate Eq. (22) after substituting for the unobserved time series D_t with functions of time (e.g., a linear time trend) and/or observable proxies for relative skill demand shifts (such as an index of between-industry demand shifts, cyclical indicators, or measures of international trade). This procedure typically involves OLS estimation of Eq. (22) using national time series data under the assumption that relative skill quantities employed are pre-determined and yields direct estimates of σ and of the impact of observable demand shifters (e.g., Freeman, 1975, 1978; Katz and Revenga, 1989). The same basic approach can be implemented on panel data on wage structure changes by regions (Juhn, 1994; Topel, 1993) or countries. The strong assumption of exogenous relative supply shifts and standard problems of estimation from time series samples with non-independent observations should introduce a note of caution in interpreting such estimates.

Katz and Murphy (1992) implement this approach to explain changes in the US college/high school wage differential from 1963 to 1987. The precise relative wage measure used is the ratio of (fixed-weighted) average wages of those with at least a college degree (16 or more years of schooling) relative to those with exactly a high school degrees (12 years of schooling). Katz and Murphy begin with 320 skill groups (defined by sex, education, and experience) and amalgamate them into two labor aggregates: college and high-school equivalents. The basic movements of these relative wage and quantity measures are summarized in Table 13 and the basic pattern of a moderate increase in the college wage premium in the 1960s, a decline in the 1970s, and a sharp increase in the 1980s is apparent in this data. Katz and Murphy assume D_t can be approximated by a simple linear time trend and estimate Eq. (22) over the 1963–1987 period by OLS yielding

$$\log(w_{ct}/w_{ht}) = \underset{(0.150)}{-0.709}\log(N_{ct}/N_{ht}) + \underset{(0.007)}{0.033}\mathit{time} + \mathit{constant}, \qquad R^2 = 0.52, \tag{23}$$

where the numbers in parentheses are standard errors.

[44] Thus this simple framework is potentially consistent with capital-skill complementarity. In this case, changes in the relative price (or supply of capital) imply shifts in D_t. For example, the nested CES aggregate production function explicitly allowing for capital-skill complementarity of Krussell et al. (1997) yields a relative wage determination equation that can be written in the same basic form as Eq. (22).

Table 13
College/high school relative wage and quantity movements, 1963–1987[a]

	Log change (multipied by 100)			
	1963–1971	1971–1979	1979–1987	1963–1987
College/high school weekly wage ratio	7.7	−10.4	12.8	10.0
Relative supply of college to high school equivalents	31.4	40.8	25.5	97.6

[a] Source: Katz and Murphy (1992, Table 8).

The actual time series of college returns and fitted values from the regression are displayed in Fig. 8. The model does a reasonable job of explaining movements in the college wage premium over this period but misses the depth of the decline from the mid to late 1970s. The implied estimate of σ, the elasticity of substitution between college and high school labor, from Eq. (23) is 1.41. The time trend coefficient multiplied by the implied estimate of σ indicates a secular shift in relative demand favoring college workers of approximately 4.6 log points a year over this period in comparison to relative supply growth of 3.9 log points year. The model implies that strong secular relative demand growth for college graduates is necessary to explain the overall rise in the college wage

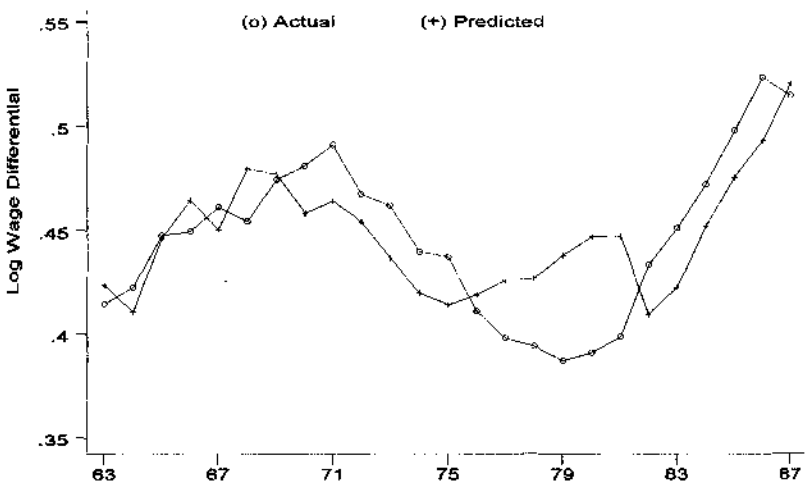

Fig. 8. Actual versus predicted log college wage premium, 1963–1987.

premium in the face of rapid relative supply growth from 1963 to 1987. But fluctuations in the rate of growth of the relative supply of college equivalents helps explain large differences across decades in the behavior of the college wage premium. The log college wage premium decreased by 1.3 log points annually from 1971 to 1979 and then increased by 1.6 log points annually from 1979 to 1987. The estimated model implies that almost half (1.36 log points per year) of the 2.9 log points per year difference in the increase in the log college wage premium in the 1980s from the 1970s is explained by a slowdown in relative supply growth with remaining 1.54 log points being accounted for by unmeasured (residual) increases in relative demand growth.

The limited time series evidence of estimates of equations of the form of Eq. (22) indicates negative effects of increases in the national relative supply of the more educated on educational wage differentials in other countries including Canada (Freeman and Needles, 1993; Murphy et al., 1998), Britain (Schmitt, 1995), Sweden (Edin and Holmlund, 1995), the Netherlands (Teulings, 1992), and South Korea (Kim and Topel, 1995). The estimates suggest (conditional on proxies for demand shifts) that a 10% increase in the relative supply of more-educated workers lowers their relative pay 3–7% in various countries implying aggregate elasticities of substitution in the 1–3 range. These findings are consistent with declining educational wage differentials throughout the OECD in the 1970s in the face of rapid supply growth of college graduates.[45] Countries that experienced at least modest increases in educational wage differentials in the 1980s – especially the United States and United Kingdom – tended to experience a decline in the rate of growth of the supply of college workers in the 1980s. Countries whose educational differentials did not expand in the 1980s – France, Germany, and the Netherlands – essentially maintained their 1970s rate of growth of supply of more-educated workers into the 1980s (Freeman and Katz, 1994; OECD, 1993). Freeman and Needles (1993) and Murphy et al. (1998) also find that the continued rapid expansion of the relative supply of college equivalents in Canada helps explain the much more modest increase in skill differentials in Canada than in the United States during the 1980s.

A controversial issue concerns the relevant relative supply measure when applying the supply and demand framework embodied in Eq. (22) in an open economy setting. The integrated equilibrium with incomplete specialization of a standard Hecksher–Olin trade model implies that national relative factor supplies only impact relative wage by changing world relative supplies (e.g., Leamer, 1996; and see the chapter by Johnson and Stafford in this volume). This essentially implies a horizontal relative demand curve at the national level. Single country time-series demonstrating negative relationships between (detrended) national relative skill supply and wage increases seem inconsistent with this prediction. This could arise if national relative supply changes are highly correlated among internationally integrated advanced economies (Berman et al., 1998). But differ-

[45] Historical evidence is also consistent with substantial effects of changes in relative skill supplies on relative wages. For example, Goldin and Katz (1995) find that the rapid expansion in secondary schooling during the "high school movement" in the United States from 1910 to 1940 was associated with a substantial narrowing of the relative earnings of white collar workers.

ences across countries in (detrended) relative supply growth also appear to be associated with differences in relative wage behavior even in such tightly linked economies as Canada and the United States. These findings suggests a focus on shifts in relative skill supplies and demand at national level may not be inappropriate. Changes in relative skill supplies in other countries may affect the price of traded goods and show up as a shift in D_t in Eq. (22). Johnson and Stafford (in this volume) provide a comprehensive discussion of deviations from the standard Hecksher–Olin model (such as differentiated products with some home bias in consumption demand and imperfect domestic factor mobility) which lead to a national relative wage determination equation consistent with this (implicitly) closed economy framework.

The second approach to assessing supply and demand stories for changes in the college wage premium is to use outside information to choose a value of σ and then use Eq. (22) and data on the time series of relative wages and quantities to impute the time series of D_t conditional on the assumed value of σ (Katz and Murphy, 1992; Johnson, 1997; Autor et al., 1998; Murphy et al., 1998). An advantage of this approach (conditional on knowledge of reasonable values for σ) is that one can draw inferences about the path of D_t without assuming full employment or the exogeneity of relative supply changes. One can also examine the sensitivity of different stories to "reasonable" choices for σ and determine whether the implied time series for D_t matches well with possible observable measures of demand shifts. Solving Eq. (22) for D_t and rearranging terms yields

$$D_t = \log(w_{ct}N_{ct}/w_{ht}N_{ht}) + (\sigma - 1)\log(w_{ct}/w_{ht}). \tag{24}$$

Changes in the log relative demand for college equivalents equals the sum of the change in the log relative wage bill and a term that depends positively (negatively) on the change in the log college wage premium when $\sigma > 1$ ($\sigma < 1$). If $\sigma = 1$ (the Cobb–Douglas case), then changes in the relative demand for college equivalents are directly given by changes in the relative wage bill.

This approach requires some knowledge of a plausible range for the elasticity of substitution between high- and low-education workers. The estimate of $\sigma = 1.41$ from Eq. (23) is in the middle of the range of 0.5–2.5 in earlier studies using cross-sectional approaches reviewed by Freeman (1986). Time series studies for different countries suggest a similar range. In an important early study, Johnson (1970) uses cross-state data for 1960 yielding estimates of the elasticity of substitution of college and high school labor of close to 1.5. Krussell et al., 1997 have extended the Katz–Murphy model of Eq. (23) through 1991 (using a slightly different aggregation scheme into college and high school workers) and find a similar implied estimate of σ of approximately 1.3. Krusell et al. generate a modestly higher estimate of $\sigma = 1.67$ from a more structural model directly allowing for capital-skill complementarity and replacing the linear time trend proxy for D_t with a measure of the relative supply of capital equipment. Heckman et al. (1998) develop a distinctive approach to measuring relative skill prices and quantities for two skill groups that allows for movements in wages to deviate from movements in skill prices because of changes in the amount of earnings potential devoted to on-the-job training. Heckman et al.

estimate the elasticity of substitution between high and low skill labor to be 1.44 by applying OLS to (22) for March CPS data from 1965 to 1990, and find quite similar estimates of σ when instrumenting for relative employment shares with cohort size. In summary much recent evidence suggests the elasticity of substitution between college and non-college workers in the United States is close to 1.4, but a substantial range of uncertainty remains.[46]

Autor et al. (1998) assess alternative explanations for changes in the US college wage premium from 1940 to 1996 under different assumptions about σ. They divide the work force into two groups: college equivalents (college graduates plus half of those with some college) and high school equivalents (half of those with some college plus workers with 12 or fewer years of schooling).[47] Panel A of Table 14 shows decadal changes in the log college/high wage differential and the log relative wage bill and supply of college equivalents. The total wage bills for college equivalents and high school equivalents can be directly calculated from household data on employment and earnings and the college/high school wage premium is estimated in each year from a standard human capital log earnings equation with individual year of schooling dummies. The (composition-adjusted) log relative supply change is calculated simply as the change in log relative wage bill minus the change in the (regression-adjusted) log relative wage: $\log(N_{ct}/N_{ht}) = \log([w_{ct}N_{ct}/w_{ht}N_{ht}]) - \log(w_{ct}/w_{ht})$. The 1970s is clearly the outlier decade in terms of the rapid relative supply growth of college graduates associated with the labor market entry of the baby boom cohorts and possible effects of incentives for college enrollment from the Vietnam War.

The sensitivity of conclusions concerning the implied time path of the growth of relative demand for college workers from (24) under different assumptions about the magnitude of σ is illustrated in panel B of Table 14. The base case assumption of $\sigma = 1.4$ implies the sharp difference in the behavior of the college wage in the 1970s and the 1980s can be attributed both to slower relative supply growth and faster relative demand growth. An acceleration in relative demand growth is necessary to explain the sharp rise in the college wage premium in the 1980s for estimates of σ in the range of most recent estimates from 1 to 2. A marked decrease in the rate of growth of relative demand is apparent in the 1990s. The compression of educational wage differentials in the 1940s is attributed to slow (and possibly negative) relative demand growth for college workers. Goldin and Margo (1992) find particularly strong demand growth for unskilled labor during the 1940s, but they also conclude that wage compression in the 1940s was at least partially driven by institutional factors including direct government intervention in wage setting during World War II, the

[46] Furthermore there is little reason to expect technological changes to leave σ relatively constant and increased openness is likely to imply greater substitutability of domestic and foreign labor and an implied increase in σ. But little direct evidence is available on changes in the aggregate elasticity of substitution.

[47] Johnson (1997) defines college equivalents in the same manner. The findings are quite similar when the more formal approach of Katz and Murphy (1992) is used to allocate different education groups to college and high school equivalents, or when a classification of workers into college graduates and those without college degrees (less than 16 years of completed schooling) is used.

Table 14
College equivalent wage-bill shares, supply and demand shifts, 1940–1996[a]

A. Changes in college-plus/non-college log relative wages, wage bill, and supply (100 × annual log changes)

	Relative wage	Relative wage bill	Relative supply change
1940–1950	−1.86	0.50	2.35
1950–1960	0.83	3.75	2.91
1960–1970	0.69	3.25	2.55
1970–1980	−0.74	4.25	4.99
1980–1990	1.51	4.05	2.53
1990–1996	0.40	2.81	2.41

B. Implied relative demand shifts favoring college-equivalents (100 × annual log changes)

	$\sigma = 1$	$\sigma = 1.4$	$\sigma = 2$
1940–1950	0.50	−0.25	−1.36
1950–1960	3.75	4.08	4.58
1960–1970	3.25	3.52	3.94
1970–1980	4.25	3.95	3.50
1980–1990	4.05	4.65	5.56
1990–1996	2.81	2.97	3.21

[a] Source: Autor et al. (1998, Table 2).

rapid expansion of unions, and possible changes in previous customary wage setting norms.

Overall Table 14 indicates rapid growth in the relative demand for college graduates since 1950 is necessary to reconcile the large increase in the US college wage premium in the face of continuing relative supply increases. Relative supply and demand fluctuations appear to play roles in decadal variations in the change in the college wage premium. The hypothesis of an acceleration in relative demand growth in the 1980s possibly from the computer revolution or globalization factors is supported assuming σ is in the range of recent estimates of 1.3–1.7. But the slowdown in demand growth in the 1990s is surprising from this perspective given the continuing spread of computers and more rapid growth of US trade with less-developed countries in the first half of the 1990s than in the 1980s (Borjas et al., 1997). Splitting the full time period roughly in half into the 1940–1970 and 1970–1996 sub-periods, there is a faster rate in the rate of relative demand growth the second half of the sample suggestive of hypotheses of an increased rate of skill-biased technological change starting in the 1970s (Greenwood and Yorukoglu, 1997). But evidence of a discrete trend break in the 1970s is not very strong.

These findings indicate the importance of assessing potential sources of trend growth in favor of more-educated workers (such as skill-biased technological changes, capital-skill complementarity, and steady increases in globalization) as well as sources of variation in the rate of demand shifts across periods and the sources of variation in the rate of supply growth (e.g., cohort size, access to higher education, immigration).

5.4. Between- and within-industry shifts in relative demand

From the late 1970s to the mid-1990s groups of workers (defined by education and other measures of skill and by sex) with rising relative wages have also tended to have rising relative supplies in most advanced nations (Katz et al., 1995; Berman et al., 1998). This pattern is suggestive of pronounced demand shifts favoring the more educated over the less educated and women over men. Substantial shifts in relative demand favoring more-educated workers appear necessary to explain wage structure changes in the United States and other OECD nations both over recent decades and probably over the past century (e.g., Tinbergen, 1974, 1975; Gottschalk and Smeeding, 1997).

Changes in product demand ("deindustrialization"), globalization factors, and skill-biased technological change have attracted much attention as possible sources for shifts in relative labor demand. A common approach is to conceptualize relative demand shifts as coming from two types of changes: those that occur within industries (i.e., shifts that change the relative factor intensities within industries at fixed relative wages) and those that occur between industries (i.e., shifts that change the allocation of total labor between industries at fixed relative wages). Sources of within-industry shifts include pure skill-biased technological change, changes in the relative prices (or supplies) of non-labor inputs (e.g., computer services or new capital equipment), and changes in outsourcing activity. Between-industry shifts in relative labor demand may be generated by sectoral differences in productivity growth and by shifts in product demand across industries arising either from domestic sources or from shifts in net international trade which change the domestic share of output in an industry at fixed wages.

This conceptualization has led to the use of decompositions of aggregate changes in the utilization of more-skilled labor into between-industry and within-industry components as a guide to the importance of product demand shifts as opposed to skill-biased technological change (or outsourcing) as sources of relative demand changes (e.g., Murphy and Welch, 1993b; Berman et al., 1994; Autor et al., 1998). Even the most detailed industry classifications available in the standard household and establishment surveys used in such analysis represent aggregates of multiple product markets. Thus, in practice, measured within-industry shifts in labor demand may contain the effects of product demand shifts within the available industry categories. This concern has motivated the use of establishment-level data to decompose changes in the overall employment share (or labor cost share) of more-skilled labor into between- and within-establishment components (e.g., Dunne et al., 1996; Bernard and Jensen, 1997). Of course, product demand shifts could potentially lead to shifts in product mix, changes in production technology, and changes in the organization of work and relative skill demands at the establishment level. Such decompositions alone clearly cannot separate out the exogenous forces driving changes in skill utilization at the plant level. These analyses should be supplemented with case studies and with attempts to examine the correlates of differences across industries and plants of the rate of skill upgrading.

The effect of between-sector shifts in labor demand on the relative demand for different

demographic (or skill) groups depends on group differences in industrial employment distributions. Shifts in employment demand between industries will have a larger effect on the relative demands for different labor inputs the greater are the differences in factor ratios (skill intensities) across industries. There exist substantial differences across industries in all advanced nations in employment distributions of different education groups and of men versus women. Changes in the industrial distribution of employment (measured in efficiency units) and variation in the utilization of highly educated (college) labor across broad US industries from 1968 to 1988 are illustrated in Table 15, which uses the college-equivalents aggregation approach of Murphy and Welch (1993b).

The table illustrates large shifts in the industrial employment distribution from 1968 to 1988 out of manufacturing sectors (especially low-skill and medium-skill manufacturing) and into professional services and finance, trade, and education and welfare services. Longer-term shifts in the industrial distribution of employment from 1940 to 1990 also show large shifts towards the more highly-educated sectors (e.g., Juhn, 1994). Industrial employment shifts since 1960 have favored industries that more intensively utilize college graduates relative to less-educated workers and women relative to men. The industries most intensive in less educated males have seen the largest decline. These patterns are reinforced when one considers occupational shifts as well industrial shifts (Katz and Murphy, 1992; Murphy and Welch, 1993a).

If within-industry relative factor demand is stable so that changes in the wage structure are entirely explained by between-industry shifts in labor demand and relative supply changes, then the shares of industrial employment of groups whose relative wages have increased should tend to fall inside every industry. Thus the hypothesis of stable within-industry demand implies that the share of college equivalents should have declined in all US industries over the past few decades. In fact, Table 15 illustrates strong within-sector upgrading occurred from 1968 to 1988 with the share of college equivalents increasing in every broad industry. Similar patterns of substantial skill upgrading are observed in the examination of changes in labor utilization within more disaggregate industries (Berman et al., 1994; Autor et al., 1998) and at the establishment level (Dunne et al., 1996; Bernard and Jensen, 1997; Doms et al., 1997).[48] This finding indicates that within-industry demand shifts favoring these groups must have occurred. On the other hand, the finding does not rule out the possibility that the between-industry shifts have also played a significant role in relative wage changes. But Murphy and Welch (1993a) and Autor et al. (1998) find that the vast majority of the increased utilization (measured by employment or labor cost) of college graduates in recent decades can be accounted for by within-industry changes. And Dunne et al. (1996) find with plant-level data for manufacturing that aggregate changes in skilled labor employment and labor cost share are dominated by within-plant changes.

How does one quantitatively assess the contributions of different sources of relative labor demand shifts? This is a difficult issue often requiring strong assumptions about

[48] Berman et al. (1994) document a similar pattern of within-industry skill upgrading (shifts to non-production workers) in the manufacturing sectors of all advanced countries in the 1970s and 1980s even during period of sharply rising relative wages for more-skilled workers.

Table 15
US employment shares and percentage college labor by industry[a]

Industry	Employment shares			Percentage college labor		
	1968	1988	Percent change 1968–1988	1968	1988	Percent change 1968–1988
Agriculture and mining	4.8	3.3	−32.5	12.7	29.4	16.7
Construction	6.8	7.1	5.1	12.5	22.2	9.8
Low-skill manufacturing	4.4	2.7	−39.5	9.4	17.1	7.7
Medium-skill manufacturing	13.0	8.0	−38.1	14.8	25.9	11.1
High-skill manufacturing	13.1	10.3	−21.6	27.7	44.8	17.0
Transportation and utilities	7.6	7.3	−3.9	15.4	34.9	19.5
Wholesale	3.9	4.6	15.4	26.7	41.8	15.0
Retail	11.8	12.2	3.5	15.7	29.9	14.2
Professional and financial	13.5	22.3	65.2	42.8	58.8	15.9
Education and welfare	9.0	10.3	13.9	73.1	75.8	2.7
Government	7.1	6.8	−4.4	30.8	50.7	19.9
Other services	5.0	5.3	5.3	12.8	27.7	14.9
All industries	100.0	100.0	0.0	26.7	43.6	16.9

[a] Source: Murphy and Welch (1993b, Table 3.4). Notes: All quantities refer to fixed-wage weighted aggregates of annual hours across experience, sex, and education. Industry shares refer to the percentage of aggregate fixed-wage weighted labor hours employed in the industry. The percentage college labor refers to the percent of fixed wage weighted labor accounted for by the college wage aggregate. See Murphy and Welch (1993b) for details of the aggregation scheme.

sectoral production functions and the consumer preferences (Bound and Johnson, 1992). One widely used measure of the effect of between-sector demand shifts on relative labor demands is the fixed-coefficient input requirements index introduced by Freeman (1975). This index measures the percentage change in the demand for a demographic group as the weighted average of percentage employment growth by industry where the weights are the industrial employment distribution for the demographic group in a base period. This proxy for the percentage change in demand for demographic group k can be written as

$$\Delta DEM_k = \sum_j \lambda_{jk}(\Delta E_j/E_j), \tag{25}$$

where j indexes industry, E_j is total employment in industry j, $\lambda_{jk} = E_{jk}/(\sum_j E_{jk})$ in a base year, and E_{jk} is the employment of group k in industry j. Katz and Murphy (1992) provide a formal justification for ΔDEM_k as a between-industry demand shift index when employment is measured in efficiency units (value-weighted labor inputs), when industry technologies are held fixed except for factor-neutral technological change, and when relative wages are unchanging. Since changes in relative wages can directly affect the distribution of industrial outputs (and employments), ΔDEM_k will not measure the effects on relative labor demand of changes in the allocation of employment across sectors at fixed wages

when relative wages are changing. These demand shifts indices will tend to understate the "true" between-industry demand shift favoring groups with rising relative wages and overstate demand shifts for groups with falling relative wages (Katz and Murphy, 1992). Murphy and Welch (1993a) and Juhn (1994) propose and implement adjustments for this bias under the strong assumption of unit own-price and zero cross-price elasticities of consumer demand.

Empirical analyses of the magnitude of between-industry and between-occupation shifts in relative labor demand using (adjusted and unadjusted) versions of ΔDEM_k indicate strong and rather steady between-industry and between-occupation demand shifts favoring more-educated workers and high-wage workers from 1950 to the present (Katz and Murphy, 1992; Juhn et al., 1993; Murphy and Welch, 1993a; Juhn, 1994). Between-industry demand shifts actually appear to be larger in magnitude in the 1960s, a period of the rapid expansion of employment in government and education-intensive service sectors, than in the period since 1970 (Katz and Murphy, 1992; Autor et al., 1998). The direction of demand shifts in the 1940s are less clear (e.g., Goldin and Margo, 1992). But the magnitudes of measured demand shifts for more-educated labor between industries or between occupations are consistently much smaller than the growth of the relative supply of more-educated workers (Katz and Murphy, 1992; Murphy and Welch, 1993a). Thus substantial within-industry and within-occupation demand shifts favoring the more-skilled are a key driving force in the large secular increase in the relative demand for more-educated workers documented in Table 14. Similar patterns are apparent in other OECD countries (e.g., Katz et al., 1995). These patterns are strongly suggestive of an important role of skill-biased technological change.[49]

When within-sector factor-biased technological changes are allowed, the interpretation of ΔDEM_k as a measure of the impact of product demand shifts on relative labor demand becomes more tenuous and the nature of the bias is more complicated (Bound and Johnson, 1992). In this case, one needs to add more structure (i.e., assumptions concerning sectoral production functions and consumer preferences) to develop measures of the contribution of product demand shifts and skill-biased technological change as sources of changes in relative labor demand. We illustrate these issues using a simplified version of the model developed by Bound and Johnson (1992) with two inputs college equivalents (c) and high school equivalents (h). Under the rather strong assumptions of Cobb–Douglas industry production functions and Cobb–Douglas consumer preferences, we find that a standard shift-share decomposition of the growth of the aggregate college wage-bill share (share of college equivalents in total costs) can be used to directly measure the extent to which the growth in the relative demand for college equivalents reflects skill-biased technological change as opposed to product demand shifts.

Following Bound and Johnson (1992), we assume the economy consists of J industries

[49] An alternative possibility for large within-industry (and within-plant) shifts in relative labor demand favoring skilled workers is increased foreign outsourcing of less-skilled jobs (Feenstra and Hanson, 1996). Berman et al. (1994, 1998) conclude that (at least through the 1980s) the amount of such foreign outsourcing is too small for it to be the driving force behind within-industry skill upgrading.

and the output of each industry j (Q_j) depends on the employment of college and high school equivalents according to a CES production function of the form of Eq. (20) with a common elasticity of substitution ($\sigma = 1/[1 - \rho]$) and with the other technology parameters (α_{jt}, a_{jt}, and b_{jt}) allowed to vary by industry and time. The relative demand for the output of industry j relative to a reference industry r in period t is assumed to be given by

$$Q_{jt}/Q_{rt} = \theta_{jt}(P_{jt})^{-\varepsilon}, \tag{26}$$

where P_{jt} is the price of Q_{jt} relative to Q_{rt} and θ_{jt} is a parameter that reflects consumer tastes and other factors (such a foreign competition) affecting relative product demand for the output of industry j in year t.

We consider the special case of a Cobb–Douglas economy: $\sigma = \varepsilon = 1$. The production function for industry j can now be written as

$$Q_{jt} = A_{jt}N_{cjt}^{\alpha_{jt}}N_{hjt}^{1-\alpha_{jt}}, \tag{27}$$

where A_{jt} indexes the level of productivity in industry j in year t. We assume the aggregate labor supplies of college equivalents (N_{ct}) and of high school equivalents (N_{ht}) are exogenous and full employment prevails so that the entire labor force of each group is allocated across the J industries: $N_{ct} = \sum_j N_{cjt}$ and $N_{ht} = \sum_j N_{hjt}$. Workers are assumed to be mobile across industries so that wages are equalized across industries. These assumptions imply (using Eq. (A8) of Bound and Johnson, 1992) that the log ratio of the competitive wage for college equivalents to that of high school equivalents is given by

$$\log\left(\frac{w_{ct}}{w_{ht}}\right) = \log\left(\frac{\sum_j \alpha_{jt}\theta_{jt}}{\sum_j (1-\alpha_{jt})\theta_{jt}}\right) - \log\left(\frac{N_{ct}}{N_{ht}}\right) = D_t - \log\left(\frac{N_{ct}}{N_{ht}}\right). \tag{28}$$

Eq. (28) is of the same form as Eq. (22) with an aggregate elasticity of substitution between college and high school equivalents of 1 and with the demand shift term D_t now directly related to industry technology and product demand shift parameters.[50]

Under these Cobb–Douglas assumptions the aggregate log relative demand for college equivalents (D_t) can be decomposed into a between-industry component that depends only on product demand shifts (changes in the θ_{jt}'s) and a within-industry component that depends only on the pace of skill-biased technological change (changes in the α_{jt}'s). These between- and within-industry demand shift components can also be directly measured with data on industry shares of the aggregate wage bill and on the share of the college wage-bill share in each industry. The Cobb–Douglas production function assumption implies that α_{jt}'s are directly measured by the share of the total wage bill accounted for by college equivalents in each industry:

$$\alpha_{jt} = (w_{ct}N_{cjt})/Y_{jt}, \tag{29}$$

[50] The common industry-level elasticity of substitution and the aggregate elasticity of substitution are only equal when $\sigma = \epsilon$ or all industries have the same factor intensities or both.

where $Y_{jt} = w_{ct}N_{cjt} + w_{ht}N_{hjt} = P_{jt}Q_{jt}$, with the last equality arising from constant returns to scale in a model with only two labor inputs. The assumption of $\varepsilon = 1$ in Eq. (26) means that the relative product demand shift for industry j ($\theta_{jt}/\sum_j \theta_{jt}$) can be directly measured by its share of aggregate revenues or by Y_{jt} (its share of the aggregate wage bill) under the normalization of $\sum_j Y_{jt} = 1$.

Differentiating the expression for D_t in Eq. (28) yields an expression for the (instantaneous) rate of change in log relative demand for college equivalents that can be written as

$$\mathrm{d}D_t = \mathrm{d}D_t^w + \mathrm{d}D_t^b, \tag{30}$$

where

$$\mathrm{d}D_t^w = \frac{\sum_j \mathrm{d}\alpha_{jt}Y_{jt}}{Y_{ct}} + \frac{\sum_j \mathrm{d}\alpha_{jt}Y_{jt}}{1 - Y_{ct}} \tag{31}$$

and

$$\mathrm{d}D_t^b = \frac{\sum_j \alpha_{jt}\mathrm{d}Y_{jt}}{Y_{ct}} + \frac{\sum_j \alpha_{jt}Y_{jt}}{1 - Y_{ct}} \tag{32}$$

and $Y_{ct} = \sum_j \alpha_{jt}Y_{jt} = w_{ct}N_{ct}/(w_{ct}N_{ct} + w_{ht}N_{ht})$, the aggregate college wage-bill share. The numerator of Eq. (31) is simply the within-industry growth component of the growth of the aggregate college-wage bill share, and the numerator of Eq. (32) is simply the between-industry component.

Thus a standard shift-share decomposition of the growth of the wage-bill (labor-cost) share of more-skilled workers (Berman et al., 1994, 1998; Autor et al., 1998) can be used to directly measure the effects of skill-biased technological change (within-industry demand growth) and product market shifts (between-industry demand growth) on overall relative demand growth. Autor et al. (1998) have implemented this approach on data for three-digit industries for the 1960 to 1996 period. They find rate of within-industry relative demand growth for college graduates appears to have increased from the 1960s to the 1970s and remained at a higher level in the 1980s and 1990s. This restrictive Cobb–Douglas framework suggests a larger impact of skill-biased technological change on the growth in the relative demand for college workers from 1970 to 1996 than in the 1960s. These results highlight the importance of more directly examining evidence on the role of skill-biased technological change in the recent widening of the wage structures of many OECD nations.

5.5. Skill-biased technological change

The deteriorating labor market outcomes of less-educated workers in most OECD economies over the past two decades despite their increasing relative scarcity strongly implies a steep decline in the relative demand for less-skilled workers. Skill-biased (or unskilled labor saving) technological change and increased exposure to international competition

from less developed countries (Stolper–Samuelson effects) have been offered as the leading candidate explanations for this demand shift.

Much indirect evidence suggests a dominant role for skill-biased technological change (associated with changes in production techniques, organizational changes, and reductions in the relative prices of computer services and new capital equipment) in the declining relative demand for the less skilled. First, as discussed in Section 5.4, the magnitude of employment (or wage bill) shifts to skill-intensive industries as measured by between-industry demand shift indices is too small to be consistent with explanations giving a leading role to product demand shifts, such as induced by greater trade with developing countries, or Hicks-neutral, sector-biased technological change. Estimates of between-industry demand shifts also show little evidence of acceleration in recent decades. Second, despite increases in the relative wages of more-skilled workers, the composition of US employment continues to shift rapidly towards more-educated workers and higher-skill occupation within detailed industries and within establishments (Berman et al., 1994; Dunne et al., 1996; Autor et al., 1998). A rise in the relative cost to firms of skilled labor should have led to within-industry and within-establishment shifts in employment towards unskilled labor in the absence of skill-biased technological change. Third, within-industry skill upgrading despite rising or stable skill premia is apparent in almost all industries in many other developed economies in the 1980s. Furthermore the cross-industry pattern of the rate of skill upgrading in manufacturing industries appears to be quite similar among advanced nations (Berman et al., 1998). These findings are consistent with an important role for *pervasive* skill-biased technological change throughout developed countries and concentrated in similar industries in each country as a major source of changes in relative skill demands. The potential impact of skill-biased technological change on the wage structure is likely to be greater the more pervasive it is across countries (Krugman, 1995; Berman et al., 1998; and see the chapter by Johnson and Stafford in this volume).[51]

More direct evidence also suggests that (broadly interpreted) skill-biased technological change is an important source of shifts in relative labor demand. Much econometric and

[51] The degree to which technological changes are pervasive across countries or localized within a single country is an important issue in assessing the likely impact on relative wages in increasingly open economies. It is the sector bias rather than the factor bias of localized technological change that determines its impact on relative wages in a small open economy operating under incomplete specialization in a standard Heckscher–Ohlin (Learner, 1996). Haskell and Slaughter (1998) provide an intriguing initial attempt to empirically examine whether differences across countries in the pattern of the sector-bias of (localized) technological change can help explain differences in changes in the relative wages of skilled workers in the 1980s. But the factor bias of technological change is often the crucial determinant of the relative wage impact in a closed economy setting. For example, the factor bias alone matters for how technological changes affect relative wages in a closed economy model with Cobb–Douglas sectoral production functions and Cobb–Douglas consumer preferences as indicated by Eq. (28). The factor bias re-emerges as an important factor in an open economy setting when technological change is pervasive across countries (since the integrated international economy as a whole can be viewed as a closed economy) and for localized technological change for a large open economy (so that the world prices of tradeables are affected by localized technological change).

case study evidence indicates that the relative utilization of more-skilled workers is positively correlated with capital intensity and the implementation of new technologies both across industries and across plants within detailed industries (e.g., Griliches, 1969; Bartel and Lichtenberg, 1987; Mark, 1987; Levy and Murnane, 1996; Doms et al., 1997). These patterns indicate that physical capital and new technologies appear to be relative complements with more-skilled workers. Thus secular increases in the capital/labor ratio could be a source of secular growth in the relative demand for skilled labor.[52] Krussell et al. (1997) present suggestive evidence that the rapid increase in the (quality-adjusted) stock of capital equipment since the early 1960s combined with strong complementarity between capital equipment and skilled labor can "account" for the trend growth in the relative demand for skills.[53]

There also appear to be strong correlations between industry-level indicators of technological change (computer investments, the growth of employee computer use, research and development (R&D) expenditures, utilization of scientists and engineers, changes in capital intensity measures) and the within-industry growth in the relative employment and labor cost share of more-skilled workers (Berndt et al., 1992; Berman et al., 1994; Wolff, 1996; Allen, 1997; Machin and Van Reenen, 1998; Autor et al., 1998). Technology indicators, particularly computer investment or employee computer usage, also appear to be more powerful explanatory variables for differences among industries in the pace of skill upgrading than are indicators of outsourcing activity, import pressures, or changes in export activity (Autor et al., 1998).[54] The causal interpretation of contemporaneous correlations of technology indicators such as R&D intensity and computer use with skill upgrading is unclear since R&D activities directly used highly-educated workers and since other sources of changes in the use of skilled workers could drive variation across industries in purchases of computers. But Autor et al. (1998), Machin and Van Reenen (1998), and Wolff (1996) find that lagged computer investments and R&D expenditures predict subsequent increases in the pace of skill upgrading. This pattern is consistent with a recent survey of US human resource managers indicating that large investments in information technology lead to changes in organizational practices that decentralize decision-making, increase worker autonomy, and increase the need for highly-educated workers (Bresnahan et al., 1998).

Plant-level studies of US manufacturing by Bernard and Jensen (1997) and Doms et al. (1997) similarly find strong positive relationships between within-plant skill upgrading and both R&D intensity and computer investments. But Doms et al. (1997) find little

[52] Such a conjecture partially motivated Griliches (1969, 1970) early seminal work on capital-skill complementarity.

[53] Their measure of the capital-skill complementarity effect on relative wages evolves similarly to a linear time trend. Thus the aggregate time series model of Krussell et al. (1997) attributes variations in changes in the skill premium around trend (such as a sharp decline in the skill premium in the 1970s and sharp rise in the 1980s) to variations in the rate of growth of the relative skill supplies and to unobserved demand shocks (the residual).

[54] But the change in export intensity does seem to have a robust positive relationship to within-industry skill upgrading even conditional on measures of computer investments (Bernard and Jensen, 1997; Autor et al., 1998).

relationship between a plant-level indicator of the number of new factory automation technologies being used and within-plant skill upgrading. In contrast, case studies by the Bureau of Labor Statistics indicate large production labor saving production innovations were adopted in the 1970s and 1980s in the electrical machinery, machinery, and printing and publishing sectors, three manufacturing industries that are among the leaders in the rate of skill upgrading in most developed countries (Mark, 1987; Berman et al., 1998).

The diffusion of computers and related technologies has attracted much attention as a possibly important measurable source of recent changes in the relative demand for skills. The share of US workers using computers on the job, an extremely crude measure of the diffusion of computer-based technologies, increased from 25% in 1984 to 47% in 1993 (Autor et al., 1998). The rapid spread of computers appears to have occurred at a similar pace in other OECD countries. For example, Card et al. (1996) report similar levels of employee computer usage in Canada, France, and the United States circa 1990. Krueger (1993) and Autor et al. (1997) document a substantial log wage premium associated with computer use (conditional on standard controls for observed worker characteristics) that increased from 0.17 in 1984 to 0.20 in 1993. The extent to which this computer wage premium represents a measure of the true returns to computer skills (the treatment effect of computer use) or largely reflects omitted characteristics of workers and their employers is a subject of much debate (see, e.g., Bell, 1996; DiNardo and Pischke, 1997). But the resolution of this debate does not directly address the issue of whether the spread of computer technologies has significantly changed organizational practices and altered relative skill demands.[55]

Computer technology may influence relative labor demand in several ways.[56] Computer business systems often involve the routinization of many white-collar tasks. Simple, repetitive tasks have proved more amenable to computerization than more complex and idiosyncratic tasks (Bresnahan, 1997). Microprocessor-based technologies have similarly facilitated the automation of many production processes in recent years. Thus direct substitution of computers for human judgement and labor is likely to have been more important in clerical and production jobs than in managerial and professional jobs. Computer-based technologies may also increase the returns to creative use of greater available information to more closely tailor products and services to customers' specific

[55] The existence of a positive computer wage differential is neither a necessary nor a sufficient condition for the diffusion of computers to have induced a shift in the relative demand for more-skilled workers and to have affected the wage structure. If computer technologies are more complementary with highly-skilled than less-skilled workers, a decline in computing costs and spread of computers could generate an increase in the relative demand for and relative wages of more-educated (and more-skilled) workers. Labor market competition could require firms both with and without computer technologies to pay equal wages to attain equally able employees. In this case a cross-section wage regression with sufficient controls for worker skills would yield no computer wage premium even though computers may have greatly raised the relative wages of the more-skilled and widened the wage structure.

[56] Bresnahan (1997) provides a descriptive theory of and illuminating historical evidence on how computers affect labor demand and organizational practices. Sichel (1997) provides a thoughtful analysis of the overall impact of the computer revolution on the US economy.

needs and to develop new products. Bresnahan (1997) posits such an organizational complementarity between computers and workers who possess both greater cognitive skills and greater "people" or "soft" skills.

The direct substitution and organizational complementarity channels both predict that an increase in the relative demand for highly-educated workers should be associated with computerization. These predictions are consistent with the findings of Autor et al. (1998) that increased computer intensity is associated with increased employment shares of managers, professionals and other highly educated workers, and with decreased employment shares of clericals, production workers, and less educated workers. Bresnahan et al. (1998) similarly find in firm-level data that greater use of information technology is associated with the employment of more-educated workers, greater investments in training, broader job responsibilities for line workers, and more decentralized decision-making.

A summary interpretation of the evidence on the impact of skill-biased technological change on recent wage structure changes is illuminated by distinguishing between two distinctive hypotheses that are sometimes confused.[57] The first is that skill-biased technological change (broadly conceived to also include capital deepening and skill-biased organizational innovations) is an important (and probably the most important) driving force behind long-run secular increases in the relative labor demand more-educated and more-skilled workers. The widespread direct evidence of capital-skill and technology-skill complementarity and indirect evidence of strong within-industry and within-plant increases in the relative demand for skill are strongly consistent with this first hypothesis. In fact, the introduction of new production technologies and increases in physical capital intensity appear to have been typically associated with increased demand for more-skilled workers throughout the 20th Century.[58]

The second hypothesis is that the impact of technological change on the relative demand for more-skilled workers *accelerated* recently (possibly in the 1980s), and this acceleration can account for the particularly large increases in wage inequality and educational wage differentials in the 1980s.

The available evidence is less definitive with respect to this hypothesis. A simply supply-and-demand analysis for the United States (such as in Table 14) indicates a particularly rapid rate of relative demand growth in the 1980s under our preferred values for the aggregate elasticity of substitution between college and non-college labor. In contrast, implied relative demand growth is much slower in the 1990s, a period of continuing rapid spread of computers. But Autor et al. (1998) find that within-industry demand growth accelerated from the 1960s to the 1970s and then stayed at this higher level through the mid-1990s. This provides some indirect evidence that the impact of skill-biased technological change on relative skill demands accelerated starting in the 1970s. Autor et al. also

[57] See Autor et al. (1998), Mishel et al. (1997b), and Gottschalk and Smeeding (1997) for further discussion of these issues.

[58] For example, Goldin and Katz (1998) show that capital-deepening, the diffusion of purchased electricity, and the introduction of continuous-process and batch methods of production greatly increased the relative demand for non-production workers and more-educated production workers in manufacturing from 1909 to 1929

provide some more direct evidence that the increase in rate of within-industry skill upgrading from the 1960s to the post-1970 period is concentrated in the most computer intensive sectors of the economy. The exceptionally rapid increase in the relative supply of college graduates in the 1970s from the labor market entry of the baby-boom cohorts delayed the impact of this demand shift on wages until the 1980s. A deceleration of relative skill supply growth from the 1970s to the 1980s and 1990s appears to be a crucial part of differences in US wage structure behavior in the 1970s and the period since the 1979.

Several conceptual issues concerning the nature of skill-biased technological change merit further consideration. One possibility is that skilled workers are more flexible and facilitate the adoption of new technologies so that all technological change increases the relative demand for more-skilled labor over some transitional period (Welch, 1970; Bartel and Lichtenberg, 1987; Greenwood and Yorukoglu, 1997). As technologies diffuse and become routinized the comparative advantage of the highly skilled declines. In this case the level of demand for skilled labor depends on the rate of innovation. Periods of large increases in the skill premium correspond to technological revolutions.[59] But an ever increasing rate of innovation seems to be necessary to generate persistent secular growth in the relative demand for more-educated workers. Furthermore the apparent slowdown in growth of the relative demand for skill in the 1990s could reflect the maturing of the computer revolution. An alternative (but potentially complementary) hypothesis is that distinctive technological innovations may have different factor biases. Some of the main technological changes of the 20th Century associated with electrification and computerization may have been skill-biased, but other innovations need not be. Mechanization in the 19th Century associated with the movement from artisanal production (intensive in skilled craft workers) to factory production (intensive in unskilled labor) appears to have been largely deskilling even though more flexible workers were likely to have been necessary to assist in the introduction of factory methods (Goldin and Katz, 1998). Under this scenario the inherent skill-biased nature of 20th Century innovations rather than an accelerating rate of innovation is the source of secular within-industry growth in the relative demand for skill.

An important further issue concerns the extent to which the rate of technological change and its direction (i.e., the extent to which technological change is skill-biased) are exogenous or are affected by changes in relative skill supplies. Acemoglu (1998), following a substantial earlier literature on induced innovation, has developed an interesting model in which increases in the proportion of skilled workers affect R&D efforts and can direct technological change in a skill-biased. Acemoglu finds it is possible for the "induced" increase in the relative demand for skills to even overshoot the increase in the relative supply of skills.

[59] Recent models of how periods of rapid technological change affect the labor market include Caselli (1997), Greenwood and Yorukoglu (1997), and Helpman and Rangel (1998).

5.6. Globalization and deindustrialization[60]

A popular culprit for rising labor market inequalities in developed countries is the increased globalization of economic activity arising from reductions in barriers to trade and reduced costs to international economic transactions. Increased trade with developing countries is commonly viewed as a driving force behind "deindustrialization" (a sharp decline in the share of employment in production jobs in manufacturing) and the woes of less-skilled workers in advanced economies (e.g., Wood, 1994, 1995, 1998). US manufacturing imports from less-developed countries (LDCs) increased from 0.8% of GNP in 1970 to 2.3% in 1980 to 2.8% in 1990 to 4.1% in 1996 (Borjas et al., 1997). Increased international capital mobility, reduced costs of international technology transfer, and greater foreign outsourcing opportunities also may increase the effective elasticity of demand facing workers in bargaining, erode their bargaining power, and reduce the extent to which internal labor markets insulate them from product market and labor market shocks (e.g., Borjas and Ramey, 1995; Rodrik, 1997; Bertrand, 1998).

A common (but controversial) method for estimating the effects of trade on labor markets is *factor content analysis* (Borjas et al., 1992, 1997; Sachs and Shatz, 1994; Wood, 1994, 1995; Lawrence, 1996). The basic approach is to determine how much of different types of labor (e.g., skilled and unskilled labor) are used to produce a country's exports, and how much would have been used in produce its imports (or the domestic goods that would have been produced in the absence of imports). The difference between the supplies of labor used in exports and imports provides an estimate of the implicit change in the relative supply of unskilled labor from trade, or, equivalently, the impact of trade on the relative demand for the unskilled. An estimate of the aggregate elasticity of substitution between skilled and unskilled labor can then be used to simulate the impact of the implicit change in relative skill supplies from trade. Increased trade will tend to have an adverse effect on less-skilled workers to the extent that import-competing industries disproportionately employ less-skilled workers and export sectors are relatively more skill-intensive. This pattern is strongly present for US trade with LDCs, but the characteristics of workers in industries with high imports and exports with other developed countries are fairly similar (Sachs and Shatz, 1994; Borjas et al., 1997).

The factor content of observed changes in net exports can provide an accurate input to assessing how changes in trade affect relative wages in limited circumstances (see the chapter by Johnson and Stafford in this volume). If one begins in autarky, then allows for trade, and trade is a modest proportion of the national economy, the change in national endowments due to the factor content of trade measures the pressure of trade for changes in relative wages (Deardorff and Staiger, 1988; Krugman, 1995). More generally, if the

[60] A comprehensive treatment of theoretical and empirical issues related to assessing the impacts of international trade on the labor market is contained in the chapter by Johnson and Stafford in this volume. The chapter by Borjas contains a detailed analysis of immigration and the wage structure. Thus we present only a brief treatment of issues concerning the role of globalization factors in recent changes in the wage structure.

changes in net exports being examined are caused by external factors (e.g., reductions in trade barriers or reductions in transportation costs, changes in factor endowments abroad), then factor content analysis may be sensible. If changes in net exports result from domestic sources (e.g., an increase in the relative supply of skilled labor leading to greater net exports of high-skill goods and lower net exports of low-skill good), then factor content analysis can be quite misleading (Leamer, 1996).

A further practical issue in factor content analysis is the how to estimate the hypothetical factor content of the domestic production that would arise to replace imports from LDCs. The standard approach is to assume LDC imports would be replaced by domestic production in the closest import-competing industry using the contemporaneous average factor proportion in the domestic import-competing industry (e.g., Sachs and Shatz, 1994). But Wood (1994, 1995) has argued persuasively that within each sector there is a wide distribution of factor proportions and labor productivity, and that LDC imports are likely to be most directly competing with the segment of an industry using the most unskilled-labor intensive production techniques. The issue is somewhat more complicated since some LDC imports may not closely compete with any domestic industry so that their absence might expand domestic demand for goods or services with quite different (and possibly even higher) skill intensities than in the assumed "import-competing" sector.

Borjas et al. (1997) examine the factor content of the growth of US trade with LDCs from 1980 to 1995. They examine the robustness of the conclusions to a wide range of assumptions concerning the factor ratios that would have been used in US industries to replace LDC imports. They find that the growth of trade with LDC's from 1980 to 1995 to a 1.4 log point increase in the implicit relative supply of high school equivalents relative to college equivalents assuming US manufactures would use the same factor ratios that prevailed in their industries in 1980 (prior to the change in LDC trade being assessed) in the absence of LDC imports. Under our preferred estimate of $\sigma = 1.4$, this implies that growth of trade with LDCs can account for only 1 log point out of a 19 log point increase in the college wage premium from 1980 to 1995. Thus demand shifts from skill-biased technological change and domestic sources of changes in relative skill supplies appear to be much more significant factors in the recent expansion of the US college wage premium than trade's impact as measured by factor contents. The impact is relatively larger if one focuses on the impact of trade on the high school dropouts. But Borjas et al. also find that increased unskilled immigration had a much larger impact on changing the implicit relative supply of the least skilled US workers than did LDC trade from 1980 to 1995.

The factor content approach may understate the effects of globalization pressures on relative wages when the threat of trade, outsourcing, or plant relocation can lead to wage changes even in the absence of new trade flows (Rodrik, 1997).[61] Borjas and Ramey (1995) explore the contribution of the erosion of industry wage differentials in trade

[61] The rate of skill-biased technological change may also be affected by globalization factors both through lower costs of technology transfer (lower cooperation costs) and through threats of foreign competition inducing "defensive innovation" (Wood, 1994, 1998).

competing durable goods manufacturing industries to increased US educational wage differentials and find it to be quite modest.

Product-price studies attempt to more directly assess the implication of the Stolper–Samuelson theorem that impacts of trade on relative wages operates through changes in the relative product price of more- and less-skill intensive. Product-price studies suffer from similar practical limitations to factor-content studies both arising from data quality issues in price data (the difficulty of separating true price from quality changes) and difficulties in trying to isolate product-price changes driven by exogenous trade-related forces rather than other sources. Slaughter (1998) provides a nice review of the emerging literature in this area and concludes that these limitations combined with a wide range of somewhat conflicting results make it difficult to draw strong conclusions from the price studies concerning the impact of international trade on wage inequality. Attempts to isolate "exogenous" international components of changes in product prices and trade flows (possibly by examining the consequences of changes in trade policy and explicit trade barriers) could be a more fruitful research strategy than standard approaches to factor content analysis and product price studies.

"Deindustrialization" (a substantial decline in manufacturing employment) is also often identified as a leading cause of poor labor market performance of less-skilled workers in advanced countries. And international trade is often viewed as the major driving force behind deindustrialization (e.g., Wood (1994, 1995, 1998). Between-industry demand shift indices (Section 5.4) do indicate that shifts out of manufacturing to more-skill intensive sectors have played some role in the decline in the relative demand for less-skilled workers. But the overall rate of between-industry demand shifts does not appear to be any larger in the period of sharp increases in wage inequality in the 1980s than in other recent decades. Nevertheless, it is striking that much of the recent increase in US wage inequality and educational wage differentials is concentrated in the period from 1979 to 1985 centered on a deep recession and containing a large appreciation of the US dollar and large decline in manufacturing employment. And the periods of extremely tight labor markets and strong demand for production workers in manufacturing during the two World Wars are the two periods of large compressions in the US wage structure during the 20th Century.[62] Furthermore studies using geographic variation across US states and metropolitan areas consistently find that larger declines in manufacturing employment are strongly positively associated (at least in the short-run) with larger increases in overall wage inequality (Juhn, 1994), residual wage inequality (Bernard and Jensen, 1998), and educational wage differentials (Borjas and Ramey, 1995; Bound and Holzer, 1997).

5.7. Summary

Supply and demand models provide a useful organizing framework for understanding

[62] The 1980s were also a period of a substantial decline in unions and erosion of the minimum wage, and the two World Wars are periods of growing union power and government intervention in the economy.

important aspects of between-group wage structure changes.[63] Supply and demand factors (the determinants of competitive wages in the SDI framework of Section 4) are important determinants of wage structure changes. Substantial secular increases in the relative demand for more-educated and more-skilled workers appear necessary to explain observed patterns of the evolution of the wage structure in developed countries over most of the last century. Shifts in the industrial and occupational distribution of employment to more skill-intensive industries and occupations can account for a significant minority of this growth in the relative demand for skills. But within-industry growth in relative labor demand favoring the more educated (within-industry skill upgrading) appears to be the major driving force in the rise in the relative demand for the more skilled. This pattern suggests a key role for skill-biased technological change in explaining relative demand shifts. Strong positive cross-industry correlations of indicators of technological change (especially indicators of the usage of computer-based technologies) and the rate of skill upgrading provides more direct evidence on the importance of skill-biased technological change. Technology factors appear to be somewhat more important than international trade changes as a source of relative demand shifts favoring the more-skilled.

Variations in the rate of growth in the relative supply of more-educated workers (college workers) appear to a be an important determinant of variations in the rate of change of educational and occupational wage differentials. Changes in cohort size, incentives for educational investments, changes in female labor force participation, and international immigration appear to be important sources of variations in relative skill supplies.[64] Detrended skill supply growth helps predict detrended changes in the college wage premium in the United States and other advanced nations. A deceleration in the rate of growth of the relative supply of college workers appears to be an important determinant of the sharp increase in US educational wage differentials in the 1980s, and especially rapid growth in relative skill supply a key determinant of the narrowing of the college wage premium in the 1970s. Countries with decelerations in relative supply growth in the 1980s are those with the largest increase in educational wage differentials.

The data are less clear on whether the recent widening of the wage structure is largely driven by an acceleration in relative demand shifts favoring the more-skilled. For the

[63] We focus on applications of supply and demand models to explaining changes in educational wage differentials in this chapter. Similar models have proved useful for examining changes in relative wages by age or experience (e.g., Freeman, 1979; Katz and Murphy, 1992; Murphy and Welch, 1992). Supply and demand models are more-difficult to apply to changes in within-group (residual) inequality that are a key component of rising US wage inequality over the last two decades. See Juhn et al. (1993) for an interesting attempt to measure between-industry and between-occupation shifts in relative demand for observed and unobserved skills based on the assumption that skills are measured by one's position (percentile) in the wage distribution.

[64] See Topel (1997) for a more thorough analysis of the impacts of alternative sources of changes in relative factor proportions. See Macunovich (1998) for an interesting and more expansive analysis of how changes in relative cohort size affect the wage distribution both through standard effects of changes in factor proportions and through changes in the level and composition of aggregate labor demand through differences over the lifecycle in consumption behavior.

United States, the pace of within-industry skill upgrading does appear to have increased since 1970, and the 1980s do appear to be a period of particularly rapid relative demand growth. But institutional factors (the erosion of unions and the minimum wage and loss of industry rates) operating in the 1980s combined with supply growth deceleration can potentially explain the observed patterns even when combined with smooth trend growth in the relative demand for more-educated workers. We next turn to an examination of how changes in labor market institutions affect the wage structure.

6. Labor market rents and labor market institutions

Large and persistent wage differentials are present across industries and establishments even after conditioning on observed measures of worker characteristics, working conditions, and non-wage employee benefits and even after controlling for (time-invariant) worker unobserved ability through individual fixed effects (e.g., Krueger and Summers, 1988; Groshen, 1991; Gibbons and Katz, 1992). Positive inter-industry wage differentials are associated with lower employee quit rates and longer queues of job applicants (Katz and Summers, 1989; Holzer et al., 1991). Thus industry and establishment wage differences appear to partially reflect variation in relative rents such as predicted by models emphasizing efficiency wage considerations and worker bargaining power (e.g., Katz, 1986; Lindbeck and Snower, 1988). Differences across countries in wage setting institutions (union and government roles in wage setting) appear to be strongly related to differences in levels of wage inequality among advanced nations especially in the lower half of the wage distribution and to differences in the magnitude of educational wage differentials (Freeman, 1993, 1996; Blau and Kahn, 1998).

The apparent importance of labor rents and institutional interventions in cross-section wage distributions suggest that these factors may also matter for changes in the distribution of wages. The same labor market shocks (e.g., from skill-biased technological change, globalization factors, or changes in skill supplies) may have different impacts on the wage structure depending on how unions and government regulations affect wage setting. Changes in labor market institutions and the incidence of labor market rents may directly lead to wage structure changes.

In this section, we first explore the role of institutional factors on recent US wage structure changes. We examine the existing research on impacts of changes in industry rents, changes in the unionization, and changes in the "bite" of the Federal minimum wage. We then briefly discuss the overall roles of supply, demand, and institutional factors in differences in wage structure changes among advanced nations. An interesting and rather unexplored topic for further research is the impact of changes in ideology and norms of fairness on wage setting (e.g., Rotemberg, 1996). The large wage structure changes in most countries during the two World Wars clearly indicate the possible importance of large shocks that change wage setting norms.

6.1. Industry rents

The large variation across industries in wages for workers with the same observed characteristics suggests that differences across groups in shifts in the industrial distribution of employment may help explain changes in the wage structure by affecting the average industry wage premium earned by different groups. The share of less-educated US employees working in high-wage durable goods manufacturing fell dramatically in the 1980s, while the share of college graduates working durable goods changed very little and the share in high-wage service industries (e.g., financial and professional services) increased substantially. Furthermore the share of female college graduates working in the low-wage education and welfare service industries declined substantially in the 1980s. These patterns are most pronounced for young workers (those with up to 9 years of potential experience) (Katz and Revenga, 1989). Changes in industry wage effects may also have differential effects across demographic and skill groups given their quite distinctive industrial employment distributions (e.g., a decline in the wage premium to construction workers has a larger effect on less-educated workers who are disproportionately employed in construction).

Much research documents that changes in the US wage structure by education, experience, and gender over the past several decades largely reflect within-industry changes rather than changes in the incidence of industry rents (e.g., Bound and Johnson, 1992; Murphy and Welch, 1993b). But changes in average industry rents do appear to have significantly contributed to widening educational wage differentials in the 1980s. For example, Murphy and Welch (1993b) find, using a 49 (approximate two-digit industry) decomposition, that the US college/high school wage differential increased 16.2% overall and 12.0% within industries. Large changes in the college wage premium occur within essentially every industry, although the changes are much more moderate in industries with large shares of public employees.[65] Thus changes in relative labor rents from differential shifts in the industrial composition of employment by education group could explain up to one-fourth of the rise in the college wage premium in the 1980s. The implied estimate should be reduced proportionately to the extent industry wage differentials represent differences in unobserved ability as opposed to "true" wage differentials from labor market rents. Bound and Johnson (1992) find similar impacts of changes in the magnitude of industry rents accruing to college and high school workers in the 1980s. The impact of a declining employment share of the less-educated in high-wage industries (durable goods manufacturing) appears to be especially important for young workers in the 1980s. Murphy and Welch (1993b) estimate that the college/high school wage differential

[65] Changes in US wage inequality and educational wage differentials in the 1980s are much smaller in the public sector than in the private sector (Katz and Krueger, 1991). These public/private differences are suggestive of the importance of how differences in wage setting institutions and political pressure on wage setting can lead to quite different relative wage responses to similar labor market shocks. The rising level of unionization in the public sector since the early 1970s as compared to substantial deunionization in the private sector may also have played a role in the smaller growth in inequality among public sector workers (Card, 1998).

increased by 26.3% for workers with 1–10 years of experience and by 20% within industries. But differences in the behavior of educational wage differentials for young workers in the 1970s and the 1980s are strikingly driven by within industry changes (changes of 33.8% overall versus 29.2% within industries). The growth of within-group (residual) wage inequality in the 1970s and 1980s is also dominated by the within-industry component (Juhn et al., 1993).[66]

The recent widening of the US wage structure also provides a potential laboratory for assessing alternative interpretations of measured inter-industry wage differentials. If industry wage differentials largely reflect differences across industries in average unobserved ability (e.g., Murphy and Topel, 1990; Abowd et al., 1998), then a sharp rise in the returns to skill should lead to a widening of measured inter-industry wage differentials in the 1980s and 1990s. Widening industry wage differentials in the 1970s (Bell and Freeman, 1991) are consistent with this hypothesis since the rise in within group inequality in the 1970s suggests a rise in the price of unobserved skills. Krueger (1998) presents a preliminary exploration of this issue for the more recent period (using data from the CPS ORG file) and finds little evidence that the dispersion of inter-industry wage differentials (the standard deviation of estimated industry wage differentials for men conditional on education and experience) increased from 1979 to 1993. Krueger finds the (adjusted) standard deviation of industry wage differentials (at the approximately two-digit level) increased sharply from 0.147 in 1979 to 0.173 in 1983 and then declined rather steadily back to 0.149 in 1993.

6.2. Unions

Unions play an important role in wage determination in all advanced nations both directly through collective bargaining and union threat (or spillover) effects on wages and indirectly by affecting government policies (e.g., minimum wages and other product and labor market regulations). Lewis (1986) concludes from a thorough review of the enormous literature on US union relative wage effects that the average treatment effect of union coverage on individual earnings (holding the locus of unionization fixed) was approximately 15% (15 log points) in the 1970s. More recent studies using longitudinal data to control for selectivity on unobserved ability into the union sector reach a similar conclusion and find a much larger union wage effect for low-skill and less-educated workers than for high-skill and more-educated workers (e.g., Card, 1996). Thus traditionally higher unionization rates among less-educated and blue-collar males are likely to have tended to serve to reduce educational and occupational wage differentials. Unions also tend to reduce wage inequality within the union sector by compressing wage differentials and standardizing wages between jobs and between establishments. Freeman and Medoff (1984) conclude for the United States that the inequality reducing effects of unions (stan-

[66] But Davis and Haltiwanger (1991) and Dunne et al. (1997) find with plant-level data that growing between-plant wage differentials are an important component of increased wage dispersion for manufacturing employees in the 1980s and early 1990s.

dardizing wages among jobs and narrowing the white collar/blue collar wage differential) have tended to be larger than the inequality increase effect of unions by creating a union/non-union wage differentials among workers who otherwise would receive similar wages.

Thus the sharp US decline in unionization over the past two decades concentrated among less-educated males could be an important source of expanding educational wage differentials and overall wage inequality for males.[67] Card (1998) estimates that the US union membership for males declined from 30.8% in 1973–1974 to 18.7% in 1993. The overall decline masks substantial differences by education. Among US males, the unionization rate fell from 1973–1974 to 1993 by 20.8 percentage points for those with less than 12 years of schooling, 14.8 percentage points for those with exactly 12 years of schooling, and actually increased slightly for college graduates.

A simplified version of the group wage determination model of Eq. (12) can be used to make a first-cut assessment of how changes in unionization affect between-group wage differentials. We assume the mean log wage for group k (Y_k) is the sum of the competitive wage for group k (Y_{kc}) and the product of the fraction of group-k workers that are unionized (U_k) and the union wage premium for group k (λ_k): $Y_k = Y_{kc} + \lambda_k U_k$. This approach ignores any impact of unions on non-union wages either through union threat effects or through spillover effects in which workers displaced by higher union wages increase the supply of workers to the non-union sector. The change in wages for group k is then given by a simplified version of Eq. (13):

$$dY_k = dY_{kc} + d\lambda_k U_k + \lambda_k dU_k. \tag{33}$$

Differences among groups in their changes in unionization rates and in changes in their union wage premia can affect their relative wages. Bound and Johnson (1992) implement this approach assuming a 15% union wage effect for all groups ($\lambda_k = 0.15$ for all k). Bound and Johnson find the unionizaton rate for male high school graduates fell by 11.5 percentage points from 1979 to 1998 as compared a decline of 2.8 percentage points for male college graduates. Under these assumptions the larger union decline for high school than college graduates accounted for a 1.3 log point expansion in the college wage premium for males from 1979 to 1988, or 8% of overall increase of 16.3 log points. Freeman (1993) does a full shift-share decomposition using Eq. (33) and allowing for differences in the union wage premium among education (and occupation) groups and over time. Freeman finds that de-unionization can explain a 1.5 log point increase in the male college wage premium from 1978 to 1988, but had a much larger impact (4 log points) on the expansion of the college wage premium for younger males (those aged 25–34).

DiNardo et al. (1996) and Card (1998) examine the effects of deunionization on overall wage inequality for US men and women, and Freeman (1993) examine the effects on male wage variance. DiNardo et al. (1996) use a semiparametric procedure to simulate the

[67] The decline in US union density began in the mid-1950s, but the 1980s are the period of most precipitous decline.

effects of changes in union density on the full distribution of wages of both men and women. The driving force in their results is the much more compressed wage distribution for non-union males than for union males. Their approach essentially attributes the differences in wage distributions by union status to the effects of unions on the wages of union workers. The impacts of nonrandom selection of workers into the union sector and of the general equilibrium effects of unionization are not explicitly considered. The key identifying assumption is that wage densities conditional on union status and observable covariates do not depend on the unionization rate. This may be a problematic assumption to the extent changes in the unionization rate affect the degree of nonrandom selection by unobservables into the union sector and have general equilibrium effects on the union and non-union wage distributions through changes in union power, union threat effects, and union spillover effects.

DiNardo et al. (1996) simulate the effect of the decline in unionization from 1979 to 1988 on the wage distribution in 1988 by reweighting the actual 1988 union and non-union wage densities using the 1979 unionization rate rather than the 1988 unionization rate (i.e., giving larger weight to the more compressed wage distribution for non-union workers). They find that the decline in unionization from 1979 to 1988 can account for 10.7% (0.021 log points) of the 0.195 log point rise in the 90–10 log wage differential for males and has almost no effect on changes in wage inequality for females. DiNardo et al.'s (1996) results suggest the decline in unionization contributed to a "declining middle" of the male wage distribution and can "explain" one-third of the increase in the 90–50 wage differential and actually partially offsets other forces towards a widening of the 50–10 differential.

Freeman (1993) attempts to estimate the effects of deunionization on the change in the variance of log earnings of US males from 1978 to 1988. He decomposes the effects of deunionization into changes in three components of the impact of unions on the variance of male log earnings: (1) the dispersion reducing effect of unionism among blue-collar union workers; (2) the dispersion increasing effect of unionism on the earnings of blue collar worker due to the union wage differential; and (3) the dispersion-reducing effect of unionism due to the union-induced reduction in the white collar/blue collar wage differential. Standard cross-section based estimates of each of these union effects are used in these calculations. Freeman concludes that the decline in union density can explain approximately 20% of the rise in male earnings inequality from 1978 to 1988 through these three mechanisms. Card (1998) generalizes Freeman's approach to account for nonrandom selection of workers into the union sector on estimates of union wage differentials and union effects on wage dispersion within the union sector. Card's adjusted estimates suggests somewhat more modest effects than those using standard cross-section estimates of union impacts. Card concludes that declining unionization can explain about 12% of the rise in male wage inequality (variance in log wages) from 1973–1974 to 1993 and essentially none of the increase for females.

In summary, the existing literature suggests both differential declines in industry rents by skill groups and the concentration of deunionization on the less-educated contributed to the enormous increase in educational wage differentials and overall male wage inequality

in the 1980s. Key outstanding issues in the assessment of the effects of deunionization on wage structure are the importance of unmeasured general equilibrium effects of unions on the wage structure and the extent to which union density changes are endogenous responses to other labor market forces. A further open question is whether one should adjust the observed changes in wage differentials used in supply and demand analyses for the effects of changes in industry rents and unionization. If these changes do not affect relative group employments (the economy moves off the labor demand curve), then the apparent acceleration of relative demand growth for college workers in the 1980s (e.g., as shown in Table 12 for $\sigma = 1.4$) might actually reflect the erosion of the relative labor rents of less educated workers.

6.3. Minimum wage

Direct government intervention in wage setting may also be a key factor in shaping the wage structure. The Federal minimum wage potentially may have significant effects in reducing wage inequality by raising wages in the lower end of the US wage distribution as well as adverse effects on the employment of low-wage workers.[68] The nominal Federal minimum wage was fixed at \$3.35 an hour from 1981 to 1990 so that the real Federal minimum wage declined throughout this period.

The minimum wage relative to the median wage declined by almost 40 log points from 1979 to 1989 (Lee (1999). Visual inspection of US wage distributions for men and women in 1979 and the late 1980s show substantial bunching around the (relatively high) minimum wage in 1979 (especially for women) and much less bunching around the relatively low minimum wage in the late 1980s. These patterns are suggestive of a substantial possible role for the erosion of the relative (and real) value of the Federal minimum wage on the widening of the lower half of the US wage distribution in the 1980s.

DiNardo et al. (1996) simulate the effects of restoring the 1988 minimum wage to its 1979 real value under the assumptions of no disemployment effects of such a 27% increase in the minimum wage and no spillovers of the minimum wage onto the distribution of wages above the minimum wage. They find that the decline in the real value of the minimum wage from 1979 to 1988 can account for most of the increase in the 50–10 log wage differential for both men and women and 17–25% of growth in the standard deviation of log hourly wages for men and 25–30% of the increase for women. The effects of the decline in the minimum wage on the college wage premium are somewhat more modest.

The interpretation of these minimum wage impacts depends on whether it is reasonable to assume a constant real minimum wage from 1979 to 1988 would imply a constant "bite" of the minimum wage. The erosion of the real and relative minimum in the 1980s could be a political response to changes in market force that reduced the relative shadow competitive wage of less-skilled workers and increased the adverse employment effects of

[68] The recent literature suggests rather modest effects of changes in the Federal minimum wage on the employment of low-wage workers (Card and Krueger, 1995).

minimum wage increases. The declining relative employment of workers with low-predicted wages in the 1980s (e.g., Juhn et al., 1993; Murphy and Topel, 1997) despite a declining minimum wage suggests other market forces were serving to reduce the labor market opportunities of low-wage workers. The strong correlations of a declining relative minimum wage with declining relative earnings of low-wage workers appear consistent with either direction of causation.

Lee (1999) attempts to address this issue by looking at cross-state differences in the impact of the Federal minimum wage given substantial differences in wage levels across US states. Lee's approach also allows for spillover effects of the minimum wage on wages up to the median of the wage distribution. He uses state panel data and finds strong effects of the minimum wage (relative to the median wage) on lower part of state wage distributions both using cross-section (between state variation) and panel data models with state and year effects. Cross-state differences in the "effective minimum wage" and observed state wage distributions are used to estimate effect of changes in the minimum wage on the wage distribution. The key identifying assumption is that the "underlying" dispersion in a state's wage distribution is orthogonal to the state's effective minimum wage. Low-wage states must not have inherently lower wage dispersion in the bottom half of the wage distribution than high-wage states for this approach to be valid (since the cross-state uniformity of the Federal minimum wage implies a higher effective minimum wage in low-wage states). Lee finds a strong relationship across states (especially in 1979) between the effective minimum wage and compression of the lower half of wage distribution, but little systematic relation with dispersion in the upper half suggesting no inherent differences in wage dispersion by state wage levels.

Lee's (1998) uses cross-state variation in the effective minimum to estimate how the effective minimum effects the lower half of state wage distributions. He finds that essentially all of the increase in the 50–10 wage differential from 1979 to 1988 is driven by the decline in the effective Federal minimum wage. Furthermore the rise in the minimum wage from 1989 to 1991 is associated with a narrowing of wage dispersion in the lower half of the wage distribution. Lee concludes that the erosion of the minimum wage can account for much of the increase in residual wage inequality in the 1980s and a modest proportion of increases in educational wage differentials. Teulings (1998) finds even larger minimum wage impacts examining differences across four US regions and allowing for minimum wage spillovers to spread throughout the wage distribution. The large magnitudes of spillover effects of the minimum wage in the studies of Lee (1998) and Teulings (1998) studies are important issues for further scrutiny as well as the possible impacts of alternative assumptions about employment effects of the minimum wage.

6.4. The SDI model and cross-country differences in wage structure changes

The pattern of demand shifts for more-skilled workers appears relatively similar in advanced nations, but not all OECD nations have experienced sharp increases in wage dispersion and educational wage differentials similar to the United States since the end of

the 1970s. Differences in the growth of relative skill supplies appear to be an important factor in cross-country differences. Decelerations in the growth in the relative supply of skills in the 1980s seem more pronounced in the countries with the largest expansions in educational wage differentials and overall wage inequality (the United States and the United Kingdom). Differences in labor market institutions among countries and changes in those institutions influenced the recent pattern of wage inequality changes among OECD countries (Freeman and Katz, 1994, 1995). Countries where unions, employer federations, and government agencies play a larger role in wage determination had smaller increases in inequality than in the United States. The comparison of Canada and the United States is instructive since the labor market shocks from technology and trade are likely to have been fairly similar. Yet differences in the pattern of relative skill supply growth (a deceleration in the United States but not in Canada) and wage setting institutions (much greater deunionization in the United States) appear to greatly account for larger increases in educational wage differentials and overall wage inequality in the United States (Freeman and Needles, 1993; DiNardo and Lemieux, 1997).

Countries with declining influences of wage setting institutions also tend to experience larger increases in wage inequality. For example, increased wage inequality appears to coincide with declining unionization in Britain in the 1980s, with Sweden's move from peak-level bargaining to more company- and industry-based settlement in the mid-1980s, with the ending of the greater government intervention in wage setting through the *scala mobile* in Italy in the early 1990s.

A key difficultly in the separation of the effect of supply and demand factors from those of institutional factors is the usual interpretation of institutional change as an outside force that affects labor market outcomes. But institutions are not immune to market forces. Shifts in supply and demand that raise relative wage differentials will reduce the strength of centralized collective bargaining and lower union influence on wage setting (e.g., Freeman and Gibbons 1995). Institutions that go strongly against market forces face a difficult task. The fact that unionization fell in most countries in the 1980s, when market forces appear to have favored greater inequality, may be no accident. Italy's dropping of the *scala mobile*, Sweden's move away from peak-level bargaining, and the 1980s' trend toward more plant- or firm-level arrangements in France partially reflect responses to a changing economic environment, not just random variations in modes of pay setting. A better understanding of the endogenous determinants of institutional changes is a crucial issue for future work on wage structure changes.

7. Conclusions

The existing research on changes in wage structures and earnings inequality suggests several directions for future research. In particular, researchers should consider the roles of changes in labor market institutions (the incidence of labor market rents) as well as changes in competitive supply and demand factors in assessing changes in the

wage structure. A key issue in such analyses that use a full supply-demand-institutions model is how to model the effects of institutions on employment rates and composition as well as on wages. And the extent to which institutional changes reflect exogenous political events as opposed to responses to market forces is also a major factor to assess in any attempt to sort out the effects of institutions from supply and demand factors.

Analyses of wage structure changes also can benefit from taking somewhat of a longer-term historical perspective then just examining the most recent decade of data. For example, an analysis focusing on US wage structure changes in the 1980s alone would conclude little effect of supply factors since groups with rising relative wages have rising relative supplies (the more-educated, older workers, women) indicating demand shifts are the driving force. An analysis of just the 1970s might find that demographic factors (the baby boom and a rising supply of college graduates) can explain rising experience differentials and narrowing educational wage differentials even with stable demand. But a consideration of a longer horizon might (e.g., the 1960s to the 1990s) actually indicates that relative supply shifts (e.g., the growth in the relative supply of college workers) actually slowed down in the 1980s and were exceptionally fast in the 1970s and that strong secular demand shifts favoring the more-educated a key element of any explanation. The importance of factors such as skill-biased technological change and globalization pressures in the 1980s and 1990s also look different when viewed through a longer-term perspective. Cross-country comparative work and differences across regions within a country may also provide useful variation in demand and supply shocks and institutional factors.

References

Abowd, John and David Card (1989), "On the covariance structure of earnings and hours changes", Econometrica 57: 411–445.

Abowd, John, Francis Kramarz and David Margolis (1998), "High wage workers and firms", Unpublished paper (Cornell University, Ithaca, NY).

Acemoglu, Daron (1998), "Why do new technologies complement skills? Directed technical change and wage inequality", Quarterly Journal of Economics 113: 1055–1089.

Allen, Steven G. (1997), "Technology and the wage structure", Unpublished paper (University of North Carolina).

Autor, David H., Lawrence F. Katz and Alan B. Krueger (1997), "Computing inequality: have computers changed the labor market?" Working paper no. 5956 (NBER, Cambridge, MA).

Autor, David H., Lawrence F. Katz and Alan B. Krueger (1998), "Computing inequality: have computers changed the labor market?" Quarterly Journal of Economics 113: 1169–1213.

Baker, Michael (1997), "Growth rate heterogeneity and the covariance structure of life-cycle earnings", Journal of Labor Economics 15: 537–579.

Baker, Michael and Gary Solon (1998), "Earnings dynamics and inequality among Canadian men, 1976–1992: evidence from longitudinal income tax records", Unpublished paper (University of Michigan).

Bartel, Ann P. and Frank Lichtenberg (1987), "The comparative advantage of educated workers in implementing new technology", Review of Economics and Statistics 49: 1–11.

Baldwin, Robert E. and Glenn G. Cain (1997), "Shifts in relative wages: the role of trade, technology and factor endowments", Working paper no. 5934 (NBER, Cambridge, MA).

Becker, Gary S. (1962), "Investment in human capital: a theoretical analysis", Journal of Political Economy 70 (supplement): 9–49.

Becker, Gary S. (1993), Human capital, 3rd edition (University of Chicago Press, Chicago, IL).

Bell, Brian D. (1996), Skill-biased technological change and wages: evidence from a longitudinal data set (Nuffield College, Oxford University, Oxford).

Bell, Linda and Richard B. Freeman (1991), "The causes of rising interindustry wage dispersion in the United States", Industrial and Labor Relations Review 44: 275–287.

Ben-Porath, Yoram (1967), "The production of human capital and the life cycle of earnings", Journal of Political Economy 75: 352–365.

Berman, Eli, John Bound and Zvi Griliches (1994), "Changes in the demand for skilled labor within U.S. manufacturing industries: evidence from the annual survey of manufactures", Quarterly Journal of Economics 109: 367–397.

Berman, Eli, John Bound and Stephen Machin (1998), "Implications of skill-biased technological change: international evidence", Quarterly Journal of Economics 113: 1245–1279.

Bernard, Andrew B. and J. Bradford Jensen (1997), "Exporters, skill upgrading and the wage gap", Journal of International Economics 42: 3–31.

Bernard, Andrew B. and J. Bradford Jensen (1998), "Understanding increasing and decreasing wage inequality", Unpublished paper (Yale University).

Berndt, Ernst R., Catherine J. Morrison and Larry S. Rosenblum (1992), "High-tech capital formation and labor composition in U.S. manufacturing industries: an exploratory analysis", Working paper no. 4010 (NBER, Cambridge, MA).

Bernstein, Jared and Lawrence Mishel (1997), "Has wage inequality stopped growing?" Monthly Labor Review 120: 3–16.

Bertrand, Marianne (1998), "From the invisible handshake to the invisible hand? how product market competition changes the employment relationship", Unpublished paper (Harvard University).

Blackburn, McKinley, David Bloom and Richard B. Freeman (1990), "The declining position of less-skilled American males", in: G. Burtless, ed., A future of lousy jobs? (The Brookings Institution, Washington, DC).

Bloom, David E. and Richard B. Freeman (1992), "The fall in private pension coverage in the United States", American Economic Review 82 (2): 539–548.

Borjas, George J., Richard B. Freeman and Lawrence F. Katz (1992), "On the labor market effects of immigration and trade", in: G. Borjas and R. Freeman, eds., Immigration and the work force (University of Chicago Press and NBER, Chicago, IL).

Borjas, George J., Richard B. Freeman and Lawrence F. Katz (1997), "How much do immigration and trade affect labor market outcomes?" Brookings Papers on Economic Activity 1: 1–90.

Borjas, George J. and Valerie Ramey (1995), "Foreign competition, market power and wage inequality", Quarterly Journal of Economics 110: 1075–1110.

Boskin, Michael, Ellen Dulberger, Robert Gordon, Zvi Griliches and Dale Jorgenson (1996), "Toward a more accurate measure of the cost of living", Final Report to the Senate Finance Committee.

Bound, John and Harry J. Holzer (1997), "Demand shifts, population adjustments and labor market outcomes during the 1980s", Unpublished paper (University of Michigan).

Bound, John and George Johnson (1991), "Wages in the United States during the 1980s and beyond", in: M. Kosters, ed., Workers and their wages (American Enterprise Institute, Washington, DC).

Bound, John and George Johnson (1992), "Changes in the structure of wages in the 1980s: an evaluation of alternative explanations", American Economic Review 82: 371–392.

Bresnahan, Timothy F. (1997), "Computerizaton and wage dispersion: an analytical reinterpretation", Unpublished paper (Stanford University).

Bresnahan, Timothy F., Erik Brynjofsson and Lorin M. Hitt (1998), "How do information technology and work place organization affect labor demand?" Unpublished paper (Stanford University).

Buchinsky, Moshe (1994), "Changes in the U.S. wage structure 1963–1987: an application of quantile regression", Econometrica 62: 405–458.

Buchinsky, Moshe and Jennifer Hunt (1996), "Wage mobility in the United States", Working paper no. 5455 (NBER, Cambridge, MA).

Card, David (1990), "Unexpected inflations, real wages and employment determination in union contracts", American Economic Review 80: 669–688.

Card, David (1996), "The effect of unions on the structure of wages: a longitudinal analysis", Econometrica 64: 957–979.

Card, David (1997), "Immigrant inflows, native outflows and the local labor market impacts of higher immigration", Working paper no. 5927 (NBER, Cambridge, MA).

Card, David (1998), "Falling union membership and rising wage inequality: what's the connection?" Working paper no. 6520 (NBER, Cambridge, MA)

Card, David, Francis Kramarz and Thomas Lemieux (1996), "Changes in the relative structure of wages and employment: a comparison of the United States, Canada and France", Working paper no. 5487 (NBER, Cambridge, MA).

Card, David and Alan B. Krueger (1995), Myth and measurement: the new economics of the minimum wage (Princeton University Press, Princeton, NJ).

Card, David and Thomas Lemieux (1996), "Wage dispersion, returns to skill and black-white wage differentials", Journal of Econometrics 74: 319–361.

Card, David and Thomas Lemieux (1999), "Can falling supply explain the rising return to college for younger men?" Unpublished paper (University of California, Berkeley, CA).

Caselli, Francesco (1997), "Technological revolutions", Unpublished paper (University of Chicago).

Chay, Kenneth Y. and David S. Lee (1996), "Changes in relative wages in the 1980s: returns to observed and unobserved skills and black-white wage differentials", Working paper no. 372 (Industrial Relations Section, Princeton University, Princeton, NJ).

Cawley, John, James J. Heckman and Edward Vytlacil (1998), "Cognitive ability and the rising return to education", Working paper no. 6388 (NBER, Cambridge, MA).

Chiswick, Carmel U. (1979), "The growth of professional occupations in U.S. manufacturing: 1900–1973", Research in Human Capital and Development 1: 191–217.

Cullen, Donald E. (1956), "The inter-industry wage structure, 1899–1950", American Economic Review 46: 353–369.

Cutler, David M. and Lawrence F. Katz (1991), "Macroeconomic performance and the disadvantaged", Brookings Papers on Economic Activity 2: 1–74.

Davis, Steven J. (1992), "Cross-country patterns of changes in relative wages", NBER Macroeconomics Annual 7: 239–300.

Davis, Steven J. and John Haltiwanger (1991), "Wage dispersion within and between manufacturing plants", Brookings Papers on Economic Activity: Microeconomics, 115–180.

Deardorff, Alan V. and Robert W. Staiger (1988), "An interpretation of the factor content of trade", Journal of International Economics 24: 93–107.

DiNardo, John, Nicole Fortin and Thomas Lemieux (1996), "Labor market institutions and the distribution of wages, 1973–1992: a semi-parametric approach", Econometrica 64: 1001–1044.

DiNardo, John and Thomas Lemieux (1997), "Changes in wage inequality in Canada and the United States: do institutions explain the difference?" Industrial and Labor Relations Review 50: 629–651.

DiNardo, John and Jörn-Steffen Pischke (1997), "The returns to computer use revisited: have pencils changed the wage structure too?" Quarterly Journal of Economics 112: 291–303.

Doms, Mark, Timothy Dunne and Kenneth R. Troske (1997), "Workers, wages and technology", Quarterly Journal of Economics 112: 253–290.

Douglas, Paul H. (1930), Real wages in the United States, 1890–1926 (Houghton Mifflin, Boston, MA).

Dunne, Timothy, John Haltiwanger and Kenneth R. Troske (1996), "Technology and jobs: secular changes and cyclical dynamics", Working paper no. 5656 (NBER, Cambridge, MA).

Dunne, Timothy, John Haltiwanger and Kenneth R. Troske (1997), "Wage and productivity dispersion in U.S. manufacturing: the role of computer investment", Unpublished paper (University of Missouri).

Edin, Per Anders and Bertil Holmlund (1995), "The Swedish wage structure: the rise and fall of solidarity wage policy?" in R. Freeman and L. Katz, eds., Differences and changes in wage structures (University of Chicago Press and NBER, Chicago, IL).

Even, William E. and David A. McPherson (1994), "Why did male pension coverage decline in the 1980s?" Industrial and Labor Relations Review 47: 439–453.

Farber, Henry S. (1986), "The analysis of union behavior", in: O. Ashenfelter and R. Layard, eds., Handbook of labor economics, Vol. 2 (North-Holland, Amsterdam).

Farber, Henry and Helen Levy (1998), "Recent trends in employer-sponsored health insurance coverage: are bad jobs getting worse?", Unpublished paper (Princeton University).

Feenstra, Robert C. and Gordon H. Hanson (1996), "Globalization, outsourcing and wage inequality", American Economic Review 86: 240–245.

Fortin, Nicole M. and Thomas Lemieux (1997), "Institutional changes and rising wage inequality: is there a connection?" Journal of Economic Perspectives 11: 75–96.

Frank, Robert H. and Phillip J. Cook (1995), The winner-take-all society (Free Press, New York).

Freeman, Richard B. (1975), "Overinvestment in college training?" Journal of Human Resources 10: 287–311.

Freeman, Richard B. (1978), "The effect of increased relative supply of college graduates on skill differences and employment opportunities", in: Z. Griliches et al., eds., Income distribution and economic inequality (Campus Verlag, Frankfurt).

Freeman, Richard B. (1979), "The effect of demographic factors on the age-earnings profile in the United States", Journal of Human Resources 14: 289–318.

Freeman, Richard B. (1986), "Demand for education", in: O. Ashenfelter and R. Layard, eds., Handbook of labor economics, Vol. 1 (North-Holland, Amsterdam).

Freeman, Richard B. (1993), "How much has de-unionization contributed to the rise in male earnings inequality?" in: S. Danziger and P. Gottschalk, eds., Uneven tides (Russell Sage, New York).

Freeman, Richard B. (1996), "Labor market institutions and earnings inequality", New England Economic Review, Special Issue (May/June): 157–168.

Freeman, Richard B. (1997), When earnings diverge (National Planning Association, Washington, DC).

Freeman, Richard B. and Robert Gibbons (1995), "Getting together and breaking apart: the decline in centralized bargaining", in: R. Freeman and L. Katz, eds., Differences and changes in wage structures (University of Chicago Press and NBER, Chicago, IL).

Freeman, Richard B. and Lawrence F. Katz (1994), "Rising wage inequality: the United States vs. other advanced countries", in: R. Freeman, ed., Working under different rules (Russell Sage Foundation, New York).

Freeman, Richard B. and Lawrence F. Katz (1995), Differences and changes in wage structures (University of Chicago Press, Chicago, IL).

Freeman, Richard B. and James L. Medoff (1984), What do unions do? (Basic Books, New York).

Freeman, Richard B. and Karen Needles (1993), "Skill differentials in Canada in an era of rising labor market inequality", in: D. Card and R. Freeman, eds., Small differences that matter (University of Chicago Press and NBER, Chicago, IL).

Gibbons, Robert and Lawrence F. Katz (1992), "Does unmeasured ability explain inter-industry wage differences?" Review of Economic Studies 59: 515–535.

Gittleman, Maury and Mary Joyce (1995), "Earnings mobility in the United States, 1967–91", Monthly Labor Review, 3–13.

Gittleman, Maury and Mary Joyce (1996), "Earnings mobility and long-run inequality: an analysis using matched CPS data", Industrial Relations 35: 180–196.

Goldin, Claudia and Robert Margo (1992), "The great compression: the wage structure in the United States at mid-century", Quarterly Journal of Economics 107: 1–34.

Goldin, Claudia and Lawrence F. Katz (1995), "The decline of non-competing groups: changes in the premium to education, 1890 to 1940", Working paper no. 5202 (NBER, Cambridge, MA).

Goldin, Claudia and Lawrence F. Katz (1998), "The origins of technology-skill complementarity", Quarterly Journal of Economics 113: 693–732.

Gottschalk, Peter (1997), "Inequality in income, growth and mobility: the basic facts", Journal of Economic Perspectives 11: 21–40.

Gottschalk, Peter and Robert Moffitt (1992), "Earnings and wage distributions in the NLS, CPS and PSID", Part 1, Final Report to the U.S. Department of Labor (Brown University).

Gottschalk, Peter and Robert Moffitt (1994), "The growth of earnings instability in the U.S. labor market", Brookings Papers on Economic Activity 2: 217–272.

Gottschalk, Peter and Robert Moffitt (1998), "Changes in job and earnings instability in the panel study of income dynamics and the survey of income and program participation", Unpublished paper (Boston College).

Gottschalk, Peter and Timothy M. Smeeding (1997), "Cross-national comparisons of earnings and income inequality", Journal of Economic Literature 35: 633–687.

Greenwood, Jeremy and Mehmet Yorukoglu (1997), "1974", Carnegie Rochester Series on Public Policy 46: 49–95.

Griliches, Zvi (1969), "Capital-skill complementarity", Review of Economics and Statistics 51: 465–468.

Griliches, Zvi (1970), "Notes on the role of education in production functions and growth accounting", in: W.L. Hansen, ed., Education, income and human capital (NBER and Columbia University Press, New York).

Groshen, Erica (1991), "Sources of intra-industry wage dispersion: how much do employers matter?" Quarterly Journal of Economics 106: 869–884.

Groshen, Erica and David I. Levine (1997), "The rise and decline(?), of U.S. internal labor markets", Unpublished paper (Federal Reserve Bank of New York).

Haider, Steven J. (1997), "Earnings instability and earnings inequality of males in the United States: 1967–1991", Unpublished paper (University of Michigan).

Hall, Brian J. and Jeffrey B. Liebman (1998), "Are CEOs paid like bureaucrats?" Quarterly Journal of Economics 113: 653–691.

Hamermesh, Daniel (1993), Labor demand (Princeton University Press, Princeton, NJ).

Hamermesh, Daniel (1999), "Changing inequality in markets for workplace amenities", Quarterly Journal of Economics 114: in press.

Haskell, Jonathan E. and Matthew J. Slaughter (1998), "Does the sector bias of skill-biased technical change explain changing wage inequality?" Unpublished paper (Dartmouth College).

Heckman, James J., Lance Lochner and Christopher Taber (1998), "Explaining rising wage inequality: explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents", Review of Economic Dynamics 1: 1–58.

Heckman, James J. and John J. Donahue (1991), "Continuous versus episodic change: the impact of civil rights policy on the economic status of blacks", Journal of Economic Literature 29: 1603–1643.

Helpman, Elhanan and A. Rangel (1998), "Adjusting to a new technology: experience and training", Unpublished paper (Harvard University).

Holzer, Harry J., Lawrence F. Katz and Alan B. Krueger (1991), "Job queues and wages", Quarterly Journal of Economics 106: 739–768.

Jackman, R., R. Layard, M. Manacorda and B. Petrongolo (1997), "European v. U.S. unemployment: different responses to increased demand for skill?" Discussion paper no. 349 (LSE Centre for Economic Performance).

Jaeger, David (1995), "Skill differences and the effect of immigrants on the wages of natives", Unpublished paper (US Bureau of Labor Statistics).

Jaeger, David A. (1997a), "Reconciling the old and new census bureau education questions: recommendations for researchers", Journal of Business and Economics Statistics 15: 300–309.

Jaeger, David A. (1997b), "Reconciling educational attainment questions in the CPS and census", Monthly Labor Review, 36–40.

Johnson, George (1970), "The demand for labor by education category", Southern Economic Journal 37: 190–204.

Johnson, George (1997), "Changes in earnings inequality: the role of demand shifts", Journal of Economic Perspectives 11: 41–54.

Johnson, George and Frank Stafford (1998), "Technology regimes and the distribution of real wages", Unpublished paper (University of Michigan).

Juhn, Chinhui (1992), "Decline of male labor market participation: the role of declining market opportunities", Quarterly Journal of Economics 107: 79–121.

Juhn, Chinhui (1994), "Wage inequality and industrial change: evidence from five decades", Working paper no. 4684 (NBER, Cambridge, MA).

Juhn, Chinhui, Dae-Il Kim and Francis Vella (1996), "Education, skills and cohort quality", Unpublished paper (University of Houston).

Juhn, Chinhui, Kevin M. Murphy and Brooks Pierce (1993), "Wage inequality and the rise in returns to skill", Journal of Political Economy 101: 410–442.

Karoly, Lynn (1993), "The trend in inequality among families, individuals and workers in the United States: a twenty-five year perspective", in: S. Danziger and P. Gottschalk, eds., Uneven tides (Russell Sage, New York).

Karoly, Lynn and Gary Burtless (1995), "Demographic change, rising earnings inequality and the distribution of well-being, 1959–1989", Demography 32: 379–405.

Katz, Lawrence F. (1986), "Efficiency wage theories: a partial evaluation", NBER Macroeconomics Annual 1: 235–276.

Katz, Lawrence F. (1994), "Active labor market policies to expand employment and opportunity", in: Reducing unemployment: current issues and options. A symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, WY.

Katz, Lawrence F. and Alan B. Krueger (1991), "Changes in the structure of wages in the public and private sectors", Research in Labor Economics 12: 137–172.

Katz, Lawrence F., Gary Loveman, David Blanchflower (1995), "A comparison of changes in the structure of wages in four OECD countries", in: R. Freeman and L. Katz, eds., Differences and changes in wage structures (University of Chicago Press and NBER, Chicago, IL).

Katz, Lawrence F. and Kevin M. Murphy (1992), "Changes in relative wages, 1963–87: supply and demand factors", Quarterly Journal of Economics 107: 35–78.

Katz, Lawrence F. and Ana L. Revenga (1989), "Changes in the structure of wages: the United States vs. Japan", Journal of the Japanese and International Economies 3: 522–553.

Katz, Lawrence F. and Lawrence H. Summers (1989), "Industry rents: evidence and implications", Brookings Papers on Economic Activity: Microeconomics, 209–275.

Kim, Dae-Il and Robert H. Topel (1995), "Labor market institutions and economic growth: lessons from Korea's industrialization, 1970–1990", in: R. Freeman and L. Katz, eds., Differences and changes in wage structures (University of Chicago Press and NBER, Chicago, IL).

Krueger, Alan B. (1993), "How computers changed the wage structure: evidence from micro data", Quarterly Journal of Economics 108: 33–60.

Krueger, Alan B. (1998), "Thoughts on globalization, unions and labor market rents", Unpublished paper (Princeton University).

Krueger, Alan B. and Lawrence H. Summers (1988), "Efficiency wages and the inter-industry wage structure", Econometrica 56: 259–294.

Krugman, Paul (1995), "Technology, trade and factor prices", Working paper no. 5355 (NBER, Cambridge, MA).

Krussell, Per, Lee E. Ohanian, José-Víctor Ríos-Rull and Giovanni L. Violante (1997), "Capital-skill complementarity: a macroeconomic analysis", Staff report no. 239 (Federal Reserve Bank of Minneapolis).

Lawrence, Robert Z. (1996), Single world, divided nations? (Brookings and OECD, Paris).

Leamer, Edward E. (1996), "In search of Stolper-Samuelson effects on U.S. wages", Working paper no. 5427 (NBER, Cambridge, MA).

Lee, David S. (1999), "Wage inequality in the U.S. during the 1980s: rising dispersion or falling minimum wage?" Quarterly Journal of Economics 114: in press.
Lerman, Robert (1997), "Reassessing trends in U.S. earnings inequality", Monthly Labor Review 120: 17–25.
Levinson, Alec (1998), "Understanding long-run trends in part-time and temporary employment", Unpublished paper (Milken Institute).
Lewis, H.G. (1986), Union relative wage effects (University of Chicago Press, Chicago, IL).
Levy, Frank and Richard J. Murnane (1992), "U.S. earnings levels and earnings inequality: a review of recent trends and proposed explanations", Journal of Economic Literature 30: 1333–1381.
Levy, Frank and Richard Murnane (1996), "With what skills are computers a complement?" American Economic Review 86 (2): 258–262.
Lindbeck, Assar and Dennis Snower (1988), The insider-outsider theory (MIT Press, Cambridge, MA).
Machin, Stephen and John Van Reenen (1998), "Technology and changes in skill structure: evidence from seven OECD countries", Quarterly Journal of Economics 113: 1215–1244.
Macunovich, Diane J. (1998), "Relative cohort size and inequality in the United States", American Economic Review 88 (2): 259–264.
Mark, Jerome S. (1987), "Technological change and employment: some results from BLS research", Monthly Labor Review 110: 26–29.
Mincer, Jacob (1974), Schooling, experience and earnings (NBER, New York).
Mincer, Jacob (1991), "Human capital, technology and the wage structure: what do time series show?" Working paper no. 3581 (NBER, Cambridge, MA).
Mishel, Lawrence, Jared Bernstein and John Schmitt (1997a), The state of working America: 1996–97 (M.E. Sharpe, Armonk, NY).
Mishel, Lawrence, Jared Bernstein and John Schmitt (1997b), "Did technology have any effect on the growth of wage inequality in the 1980s and 1990s?" Unpublished paper (Economic Policy Institute).
Moffitt, Robert and Peter Gottschalk (1995), "Trends in the autocovariance structure of earnings in the U.S.: 1969–1987", Unpublished paper (Boston College).
Moulton, Brent (1997), "Bias in the consumer price index: what is the evidence?" Journal of Economic Perspective 10: 169–177.
Murphy, Kevin M., W. Craig Riddell and Paul M. Romer (1998), "Wages, skills and technology in the United States and Canada", in: E. Helpman, ed., General purpose technologies (MIT Press) in press.
Murphy, Kevin M. and Robert H. Topel (1990), "Efficiency wages reconsidered: theory and evidence", in: Y. Weiss and G. Fishelson, eds., Advances in the theory and measurement of unemployment (MacMillan, London).
Murphy, Kevin M. and Robert H. Topel (1997), "Unemployment and nonemployment", American Economic Review 87 (May): 295–300.
Murphy, Kevin M. and Finis Welch (1992), "The structure of wages", Quarterly Journal of Economics 107: 285–326.
Murphy, Kevin M. and Finis Welch (1993a), "Occupational change and the demand for skill, 1940–1990", American Economic Review 83 (May): 122–126.
Murphy, Kevin M. and Finis Welch (1993b), "Industrial change and the rising importance of skill", in: S. Danziger and P. Gottschalk, eds., Uneven tides (Russell Sage, New York).
Murphy, Kevin M. and Finis Welch (1997), "The structure of wages revisited", Working paper no. 9724 (Private Enterprise Research Center, Texas A&M University).
Nickell, Stephen and Brian Bell (1995), "The collapse in demand for the unskilled and unemployment across the OECD", Oxford Review of Economic Policy, 40–62.
Ober, Harry (1948), "Occupational wage differentials, 1907–1947", Monthly Labor Review, August, 127–134.
OECD (1993), Employment outlook, July 1993 (OECD, Paris).
OECD (1996), Employment outlook, July 1996 (OECD, Paris).
OECD (1997), Employment outlook, July 1997 (OECD, Paris).
Phelps Brown, E.H. (1977), The inequality of pay (Oxford University Press, Oxford, UK).

Pierce, Brooks (1997), "Compensation inequality", Unpublished paper (U.S. Bureau of Labor Statistics).
Polivka, Anne (1996), "Data watch: the redesigned current population survey", Journal of Economic Perspectives 10: 169–180.
Rodrik, Dani (1997), Has international integration gone too far? (Institute for International Economics, Washington, DC).
Rosen, Sherwin (1981), "The economics of superstars", American Economic Review 71: 845–858.
Rotemberg, Julio (1996), "Perceptions of equity and the distribution of income", Working paper no. 5624 (NBER, Cambridge, MA).
Sachs, Jeffrey D. and Howard J. Shatz (1994), "Trade and jobs in U.S. manufacturing", Brookings Papers on Economic Activity (1): 1–84.
Schmitt, John (1995), "The changing structure of male earnings in Britain, 1974–1988", in: R. Freeman and L. Katz, eds., Differences and changes in wage structures (University of Chicago Press and NBER, Chicago, IL).
Sichel, Daniel E. (1997), The computer revolution (The Brookings Institution, Washington, DC).
Slaughter, Matthew J. (1998), "What are the results of product–price studies and what can we learn from their differences?" Working paper no. 6591 (NBER, Cambridge, MA).
Slichter, Sumner (1950), "Notes on the structure of wages", Review of Economics and Statistics 32: 80–91.
Tinbergen, Jan (1974), "Substitution of graduate by other labor", Kyklos 27: 217–226.
Tinbergen, Jan (1975), Income differences: recent research (North Holland, Amsterdam).
Topel, Robert H. (1993), "Regional labor markets and the determinants of wage inequality", American Economic Review 83 (2): 110–115.
Topel, Robert H. (1997), "Factor proportions and relative wages: the supply-side determinants of wage inequality", Journal of Economic Perspectives 11 (2): 55–74.
Teulings, Coen (1992), "The wage distribution in a model of matching between skills and jobs", Unpublished paper (University of Amsterdam).
Teulings, Coen (1997), "Aggregation bias in elasticities of substitution", Unpublished paper (University of Amsterdam).
Teulings, Coen (1998), "The contribution of minimum wages to increasing wage inequality: a semi-parametric approach", Unpublished paper (University of Amsterdam).
U.S. Department of Labor (1995), Report on the American workforce (US GPO, Washington, DC).
Welch, Finis (1969), "Linear synthesis of skill distribution", Journal of Human Resources 4: 312–327.
Welch, Finis (1970), "Education in production", Journal of Political Economy 78: 35–59.
Welch, Finis (1979), "Effects of cohort size on earnings: the baby boom babies' financial bust", Journal of Political Economy 87: S65–S97.
Williamson, Jeffrey G. and Peter H. Lindert (1980), American inequality: a macroeconomic history (Academic Press, New York).
Willis, Robert J. (1986), "Wage determinants: a survey and reinterpretation of human capital earnings functions", in: O. Ashenfelter and R. Layard, eds., Handbook of labor economics, Vol. 1 (North-Holland, Amsterdam).
Wolff, Edward (1996), "The growth of information workers in the U.S. economy, 1950–1990: the role of technological change", Unpublished paper (New York University).
Wood, Adrian (1994), North-south trade, employment and inequality (Clarendon Press, Oxford, UK).
Wood, Adrian (1995), "How trade hurt unskilled workers", Journal of Economic Perspectives 9: 57–80.
Wood, Adrian (1998), "Globalisation and the rise in labour market inequalities", Unpublished paper (Institute of Development Studies, University of Sussex).

Chapter 27

LABOR SUPPLY: A REVIEW OF ALTERNATIVE APPROACHES

RICHARD BLUNDELL*

University College London and Institute for Fiscal Studies

THOMAS MACURDY*

Department of Economics and The Hoover Institution, Stanford University

Contents

Abstracts 1560
JEL codes 1560
1 Introduction 1560
2 How have tax and welfare policies changed? 1563
2.1 US tax and welfare programs 1564
2.2 UK tax and welfare programs 1569
3 Recent empirical trends 1572
3.1 Data sources 1574
3.2 Participation 1577
3.3 Hours of work 1580
3.4 Real wages 1584
4 A framework for understanding labor supply 1586
4.1 The static labor supply model 1587
4.2 Multiperiod models of labor supply under certainty 1591
4.3 Multiperiod models of labor supply under uncertainty 1596
4.4 Basic empirical specifications 1598
4.5 Which elasticities for policy evaluation? 1603
5 Policy reforms and the natural experiment approach 1607
5.1 The natural-experiment approach and the difference-in-differences estimator 1608
5.2 Does the difference-in differences estimator measure behavioral responses? 1613
5.3 A review of some empirical applications 1615
6 Estimation with non-participation and non-linear budget constraints 1617
6.1 Basic economic model with taxes 1618

* We would like to thank John Pencavel for providing the US data, Christian Dustmann for the German data, Howard Reed for the British data, and Lennart Flood for the Swedish data. Jed DeVaro and Mika Kuismanen provided able research assistance and many helpful comments. We also thank Soren Blomquist, James Heckman, Ian Walker and Valerie Lechene for comments on sections of earlier drafts. Blundell thanks the ESRC Centre for the Microeconomic Analysis of Fiscal Policy at IFS for financial support. MaCurdy gratefully acknowledges research support from NIH grant HD32055-02. Opinions expressed in this chapter are those of the authors and do not represent the official position or policy of any agency funding this research.

Handbook of Labor Economics, Volume 3. Edited by O. Ashenfelter and D. Card

6.2 Instrumental-variable estimation 1622
6.3 Maximum likelihood: convex differential constraints with full participation 1626
6.4 Maximum likelihood: convex piecewise-linear constraints with full participation 1629
6.5 Maximum likelihood: accounting for fixed costs of participation and missing wages 1635
6.6 Welfare participation and non-convex budget constraints 1638
6.7 An approach for computational simplification and discrete hours choices 1643
6.8 Survey of empirical findings for non-linear budget constraints models 1644
7 Family labor supply 1657
7.1 The basic economic model of family labor supply 1657
7.2 The collective model of family labor supply 1661
7.3 Some empirical findings for the family labor supply model 1665
8 Structural dynamic models 1672
8.1 The standard intertemporal labor supply model with participation 1672
8.2 Learning by doing and human capital 1676
8.3 Habit persistence 1680
8.4 Review of empirical results for structural dynamic models 1680
9 Closing comments 1684
Appendix A. Specifications of within-period preferences 1686
References 1689

Abstract

This chapter surveys existing approaches to modeling labor supply and identifies important gaps in the literature that could be addressed in future research. The discussion begins with a look at recent policy reforms and labor market facts that motivate the study of labor supply. The analysis then presents a unifying framework that allows alternative empirical formulations of the labor supply model to be compared and their resulting elasticities to be interpreted. This is followed by critical reviews of alternative approaches to labor-supply modeling. The first review assesses the difference-in-differences approach and its relationship to natural experiments. The second analyzes estimation with non-linear budget constraints and welfare-program participation. The third appraises developments of family labor-supply models including both the standard unitary and collective labor-supply formulations. The fourth briefly explores dynamic extensions of the labor supply model, characterizing how participation decisions, learning-by-doing, human capital accumulation and habit formation affect the analysis of the lifecycle model. At the end of each of the four broad reviews, we summarize a selection of the recent empirical findings. The concluding section asks whether the developments reviewed in this chapter place us in a better position to answer the policy-reform questions and to interpret the trends in participation and hours with which we began this review.

JEL codes: J21; J22; J24; C21; C24

1. Introduction

Consistent with its tradition, research on labor supply during the past decade has been at the forefront of developments in empirical microeconomics. At the same time, an important component of this research has rebuffed sophisticated estimation approaches in favor

of simple methods for evaluating behavioral responses underlying hours-of-work decisions. The attention devoted to the study of labor supply arises from intense interests in assessing the consequences of a wide array of public policies, ranging from tax and welfare programs to the alteration of institutional features of labor markets. A further motivation concerns the curiosity of economists in explaining the factors underlying the dramatic changes in employment patterns that have occurred in recent years, trends that show no evidence of stabilizing in the near future.

After presenting a brief overview of the phenomena stimulating recent analyses of labor supply, this chapter pursues its main purpose of reviewing the empirical developments and findings produced by this research. It focuses on work done since the surveys of Pencavel (1986) and Killingsworth and Heckman (1986), which ably summarized the labor supply literature in the previous *Handbook of Labor Economics*. We draw widely on existing research in the labor supply literature. Our discussion of methodological developments presents simplified examples to highlight essential ideas, not attempting to attribute each development to specific authors and, thus, omitting most references in this discussion. We do not claim originality in this survey, and our discussion of applications refers to many of the studies that have made the major contributions to this research area since the earlier *Handbook* surveys. It is inevitable that we have omitted references and we apologize for such omissions.

The influence of governmental programs on people's employment and hours of work is often a critical consideration in the design of policies. Indeed, the primary objective of many recent reforms in both tax and welfare programs in North America, the UK, Scandinavia and other parts of Europe has been to encourage participants to increase their work effort. Few decades match the most recent in terms of how much change has occurred in tax and welfare policies. Understanding labor supply behavior is vital in formulating proposals that build in work incentives while providing income support.

This chapter begins with a cursory description of how tax and welfare policies have changed in recent years, considering how these changes enter the picture of labor supply and its empirical analyses. For this discussion, we focus on reforms in the US and the UK. This is not simply because of our own local knowledge but also because these two countries have been at the forefront of introducing welfare and tax reforms designed to encourage work effort – in particular, the move toward in-work benefits. These "Welfare to Work" proposals form a particularly attractive background against which to motivate labor supply analysis as they are generally reforms directly aimed at addressing the decline in participation among certain types of workers. The analysis of participation in work is key to the evaluation of welfare-to-work reforms and this is the margin over which labor supply responses may be most responsive. However, to properly evaluate the impact of a welfare-to-work policy reform, such as the Earned Income Tax Credit in the US, requires a careful examination of the balance between the labor supply decisions of those individuals already working who may now face a higher benefit (or credit) reduction rate and the labor supply decisions of those individuals who may be induced to enter by such a reform. We

provide a detailed analysis of how recent policy reforms in the US and UK have changed the shape of the budget constraint facing many workers.

Any analysis of labor supply requires an understanding of the background changes in wages, participation and hours of work. In Section 3, we provide an analysis of labor supply facts highlighting the important changes in labor market participation and in the dispersion of wages. As a comparison with the US and the UK, we document the changes in these aspects of labor supply for two additional countries: Germany and Sweden. It is these changes in participation and working hours that labor supply models attempt to explain. The success of labor supply models will be judged largely in terms of their ability to explain and enhance our understanding of the changes in participation and hours.

Having motivated our analysis of labor supply with important policy questions and labor supply facts, our aim in the remaining sections is to present a comprehensive evaluation of alternative approaches to modeling labor supply. This seeks to achieve three broad objectives: to make different studies comparable by providing a unifying framework by which the results of each can be interpreted; to provide a description of the mechanics of implementing each approach and the data and assumptions required; and to identify gaps in our knowledge which can motivate future research. We have attempted to review the state of empirical knowledge on labor supply responses, and we end each section with a discussion of relevant empirical results.

The unifying framework we develop in Section 4 is designed to compare across alternative basic labor supply specifications. It should be noted at the outset that individual labor supply responses may be reflected in the choice of hours across firms rather than within any establishment. Complexities that arise from non-linear taxation, fixed costs, welfare programs, dynamics, etc. are taken up in detail in the following sections. A simple multi-period framework is used to compare across alternative static formulations, two-stage budgeting models, the Frisch model and fully-specified lifecycle models. The aim is not to dictate a single approach to estimation, but rather to evaluate precisely what can be learned from different datasets and different approaches to estimation. The wage coefficient in each approach is related to alternative elasticity measures and we ask which measure is appropriate for the evaluation of policy reforms. Even the simplest tax reform typically involves an unanticipated shift in the profile of wages. None of the standard elasticity measures fully reflects responses to such a shift and Section 4 precisely documents what is required to answer such policy questions.

Sections 5–8 consider alternative aspects and approaches to labor supply that have been adopted in the literature. We begin with a review of the application of difference in differences and natural experiments in labor supply estimation. Our aim here is to emphasize the structural assumptions underlying this approach and to relate the estimated parameters to those needed for policy analysis. A number of influential studies that have used this approach, and related approaches, are then reviewed.

Procedures by which a researcher can fully account for non-linear taxation, fixed costs, welfare participation and missing wages in estimation and simulation motivate the discus-

sion in Section 6. Again, the emphasis here is to lay out the precise assumptions and restrictions placed on behavior by alternative models. The practical issue of how to account for multiple program participation and the interactions between the tax and benefit system are highlighted. The empirical literature in this area is vast. This aspect of labor supply continues to attract considerable research interest, reflecting the recurring importance placed on the labor supply responses to tax and benefit reforms.

Placing the labor supply problem in a context where there is potentially more than one supplier of labor in the household is covered in Section 7, which reflects two important developments in this area. The first is to acknowledge the complex set of incentives faced by multiple workers once the full tax and welfare system is accounted for. The second is the introduction of alternative models of labor supply decision-making when multiple workers are located in the same family. These alternative models that seek to account for collective choices that are solutions to bargaining within the family are still in their infancy as far as empirical application are concerned. However, we are able to compare them to the standard "unitary" model and review the empirical literature that has developed to date.

Our review of alternative formulations of the labor supply model is completed in Section 8 with a discussion of dynamic models. Here we highlight generalizations of the basic multiperiod model described in Section 4 that allow for human capital and non-participation. The first-order conditions for the standard multiperiod model can be severely distorted in the presence of human capital choices. Human capital choices, or purely exogenous learning by doing, can break the separability of the intertemporal decision rule that allows simple Frisch and two-stage budgeting formulations. This is also shared by models that allow for habits. We describe the appropriate adaptation of the multiperiod model to cover these extensions and review the results from the empirical literature. We also consider the complications that arise in these models once non-participation and fixed costs are allowed for. We evaluate the trade-off between realism and computational tractability and set up the standard discrete dynamic programming formulation for this problem.

In Section 9, we conclude this chapter with a brief assessment of what has been achieved by recent research on labor supply and ask whether we are now in a better position to answer the policy-reform questions raised in Section 2 and better able to understand the labor supply facts described in Section 3. We document a large number of significant contributions across a wide range of labor supply issues but we also identify significant gaps in our knowledge which will continue to place research on labor supply at the forefront of research in labor economics for some time to come.

2. How have tax and welfare policies changed?

In few decades have we seen the marked changes in tax and welfare policies that have occurred since the early 1980s. In the US, the number of tax brackets sharply diminished with the passage of the federal tax reform in 1986. In the UK, the number and level of

higher brackets were reduced following the 1979 move away from direct taxation and towards indirect taxation. Sweden and other European countries subsequently followed this direction in reforming their income tax systems during the late 1980s and early 1990s. In both the US and the UK, in-work benefits increasingly became the main platform for encouraging low-income families to increase their work effort and incomes. In the US, the earned income tax credit (EITC) was greatly enhanced in 1993, while in the UK the Family Credit (FC) system, based on a minimum number of weekly hours worked, reduced the limit in 1992 from 24 to 16 h per week and significantly increased the number of recipients.

In 1996, the US adopted sweeping reforms in its welfare systems, all designed to induce recipients to support themselves through work. In the UK, the Family Credit system was extended to incorporate a 30-hour benefit supplement. In the 1998 budget, Family Credit was made more generous and was renamed Working Families Tax Credit (WFTC) to signify that payments would be paid through the tax system. The motivation of much research on labor supply is to predict the consequences of such reforms for hours of work and earnings. Researchers often devote considerable attention to modeling the institutional features of tax and transfer policies. This section briefly summarizes the changes that have occurred during the last decade in tax and welfare policies. We focus on policy changes for the US and the UK. The following sections explain how labor supply analyses have exploited these changes to assess their impacts on work behavior.

2.1. US tax and welfare programs

Perhaps the easiest way to convey the complexities introduced by the US tax and welfare system is to describe the number of programs in which individuals participate when they work. Workers must pay federal income taxes which account for an array of deductions, social security tax, state income tax and a variety of health and insurance taxes. If a worker's family has sufficiently low income, it may be eligible for benefits from a patchwork of different programs. These public assistance programs provide support in the form of cash income as well as in-kind support for necessities such as food, housing, medical care or home heating. The six major programs that offer the core of resource support for poor families in the US are: Aid to Families with Dependent Children (AFDC), Food Stamp Program (FSP), Supplemental Security Income (SSI), Housing Assistance, Medicaid, and the Earned Income Tax Credit (EITC). AFDC, SSI, and EITC pay cash assistance to low-income families. FSP provides food vouchers denominated in dollars to low-income households. Housing assistance programs come in two varieties: rent subsidies for occupancy of private dwellings, and low-income public housing which is built, managed and maintained by government agencies. Finally, Medicaid is an in-kind benefits program providing medical assistance to poor persons.

Describing how all of these programs have changed individually during the past decade would occupy many papers, yet this exercise would still fall short of characterizing how these policy alterations influence labor supply, as the most profound and

disconcerting effects occur when families simultaneously participate in multiple programs. Each program has its own benefit reduction rate which determines how much benefits decline as earnings increase. These rates act as tax rates on earnings, in that they dictate how much families get to keep out of any incremental earnings they receive while collecting benefits. Because benefit reduction rates are independent across programs, the combined benefit reduction rate that results when a family participates in several programs rises to staggeringly high levels that no policymaker ever intended. This, in turn, produces significant disincentives for families to work. The relevant factor in assessing the impact of these policies on labor supply is the combined effect of these programs through time.

2.1.1. How do programs in the US combine to tax earnings and provide income support? Fig. 1 shows how net governmental transfers change as a family's earnings rise, given participation in various combinations of public assistance programs. The figure depicts three scenarios: the lower curve indicates transfers when the family receives benefits from just EITC; the middle curve gives the total benefits received when the family collects food stamps in conjunction with EITC; and the upper curve measures the total transfers when the family participates in the AFDC program as well. The curves are for a single-parent family with two children living in California – only the AFDC benefit schedule depends on California residency. Other than the social security tax (about 7.5%), families at the low income level pay no federal or state income taxes. As earnings increase (i.e., moving left to right in the figure), net transfers initially rise due to the increase in EITC, regardless of the

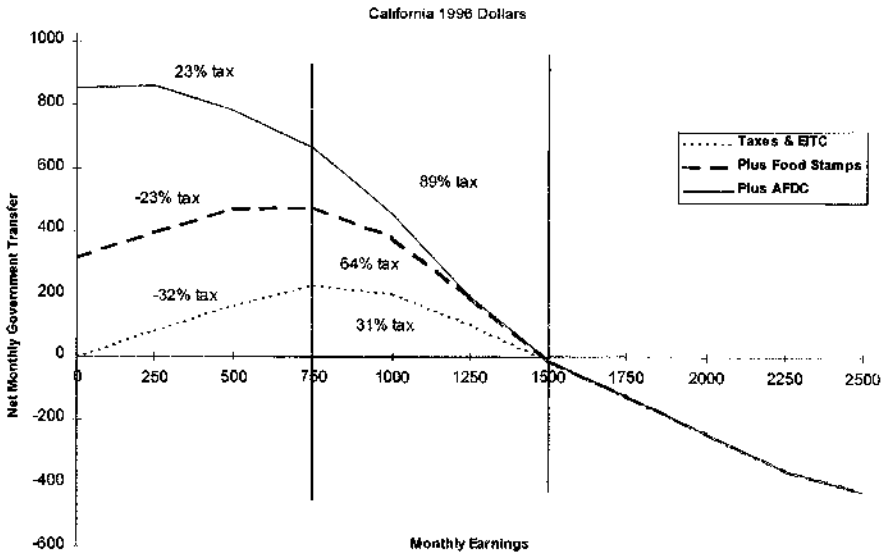

Fig. 1. Net transfers/taxes for California in 1996.

combination of programs in which the family participates. However, eventually these transfers decline with higher levels of earnings. The reversal is fastest when the family collects AFDC, food stamps and EITC simultaneously, and slowest when collecting only EITC.

For a family participating in all three programs, the uncoordinated nature of the programs leads to some unintended and undesirable features. As the family's earnings rise within the first \$750/month earned (=30 h per week at \$5.75) in 1996, the EITC provides a tax credit increasing the value of work by 40%. If this were the only program, the family would face an implicit tax rate of −32% (a negative tax), paying only social security taxes. However, since both food stamps and AFDC benefits decline more rapidly with earnings than EITC rises, a family who also participates in these programs ends up losing about 23 cents out of every \$1 earned up to \$750/month. This translates into an effective positive tax rate of 23% on earnings. Earning \$750/month, this family still receives benefits from all three programs. Increasing family earnings from \$750 to \$1500/month would put it in an income range with effective tax rates of about 89%, meaning that it would retain only 11 cents out of every dollar earned.

Ironically, this high tax rate is the result of changes during the past five years that were designed to increase work incentives. Recent federal legislation increased the generosity of the EITC, and at about the same time California lowered the benefit reduction rates through the passage of "30 and a third" reforms in AFDC. Comparing the benefit structure and tax rates in 1996 to those in 1992 reveals that these federal and California state changes decreased the effective taxes for families in the lowest earnings range. The marginal tax rate for the first \$750 of earnings fell from 71% to 23%. However, these changes simultaneously raised the marginal tax rate for the second \$750 of earnings from 59% to 89%.

Knowing that AFDC participants do not work extensively under the current system says little about their motivation or prospects for working, because the existing benefit structure creates strong disincentives to working. It is quite rational for AFDC recipients to work little or not at all. The current rules tax income highly as earnings increase. These work disincentives become more severe the more a recipient works and the closer he or she gets to self-sufficiency.

Under the system today, an AFDC recipient would need to work 40 h per week at \$6.90/h to make enough to leave AFDC (=\$1104/month). She would need to earn \$7.88/h to lose food stamps as well (=\$1261/month). Yet in moving from \$750/month to \$1500/month, her net income would rise by only \$82 due to a combination of benefit reductions in both AFDC and food stamps and a reduction in the EITC as earnings enter a "phase out range." Unfortunately, the resulting 89% tax rate falls precisely on the earnings range that makes the difference between welfare receipt and self-sufficiency.

2.1.2. How do programs differ across states?

Fig. 2 illustrates how differing AFDC programs across states affect the benefit amounts

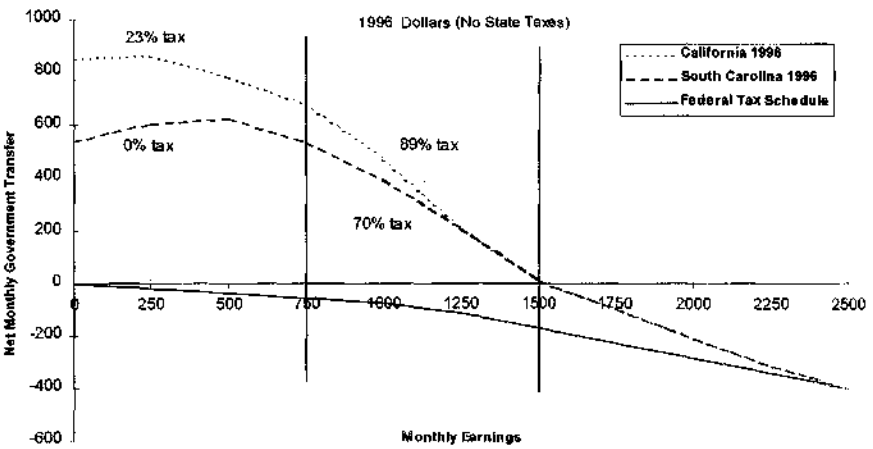

Fig. 2. Comparison of net transfers/taxes for California and South Carolina in 1996.

received by a family participating in all three programs. The top curve is for our California family and the middle curve is for an identical family living in South Carolina. We select South Carolina as the comparison state for California because, in the early 1990s, it occupied an opposite position in the distribution of state AFDC benefit levels: whereas California had the fifth most generous state AFDC program, South Carolina had the fifth least generous. The lowest curve corresponds to the taxes a family would pay if it participated in no low-income transfer programs.

Since South Carolina paid lower AFDC benefits than those in California, the net transfers received by the South Carolina family are everywhere below those of the California family until monthly earnings reach between $1250 and $1500 when both AFDC programs cease to pay benefits. The higher generosity of California's program has a serious downside: California's implicit tax rates on earnings are much higher. The benefit reduction rates are similar, but more is lost for every dollar earned in California because the reduction rate applies to a larger benefit amount.

Reduction rates are still quite large for South Carolina residents. Even though the tax rate faced by a South Carolina family increasing its earnings from $750 to $1500/month is almost 20 percentage points below the rate faced by a California family with the same earnings increase, this lower rate is still 70%. Such tax rates are staggeringly high and are very likely to discourage work.

2.1.3. How have programs changed in the US?

Fig. 3 shows how net governmental transfers have changed in California during 1985 - 1996. Similar changes have occurred in other states. These changes reflect a combination of factors, the most prominent being decreased benefit reduction rates for welfare programs and increased generosity in the EITC.

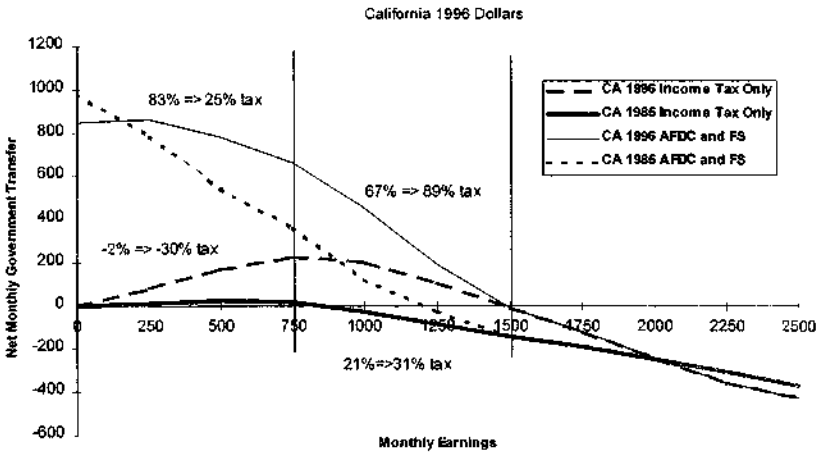

Fig. 3. Net transfers/taxes for California in 1996 versus 1985.

The two lines that begin lowest on the graph represent the net benefit receipt from EITC, combined with taxes, alone. The difference between 1985 and 1996 is striking. In 1985, the peak EITC benefit was a mere \$20, while in 1996 this figure had increased more than tenfold to \$224. As a result, the average tax rate on the first \$750 for a family receiving just EITC fell from -2% in 1985 to -30% in 1996. This reduction in tax rates leads many to argue that EITC has a strong pro-work effect. However, after \$750 EITC benefits decline, yielding a tax rate of 31% in 1996 versus 21% in 1985. Hence, the increased generosity of EITC has led to increased marginal tax rates for families seeking to increase their income from welfare-dependent levels to more self-sufficient levels, as noted above.

When combined with changes in AFDC and Food Stamps, this shift in incentives is even more pronounced. The top two lines in Fig. 3 depict total AFDC, Food Stamp and EITC benefit levels in 1985 and 1996. From 1985 to 1996, the monthly AFDC and Food Stamp benefit for a family with no earnings was reduced from \$980 to \$852. Associated with this reduction was a flattening of the benefit versus earnings graph, as shown in the figure. As a result, the combined tax rate fell from 83% to 25% on the first \$750 in earnings during this period. However, a flatter benefit reduction schedule for the first \$750/month simply required a steeper schedule for the next \$750. As a result, the average tax rate for the second \$750 in income/month – as noted, that income required to move off welfare – rose from 67% to a staggering 89% from 1985 to 1996.

In summary, changes from 1985 to 1996, which were advanced as measures to increase work incentives created markedly higher tax rates on income between \$750 and \$1500/month. This range of income is very important for families seeking to move off welfare. While it may be possible to enact further reforms that push this region of high benefit reductions and marginal tax rates to higher incomes, this can only be done at the cost of substantially reduced benefit levels or substantially increased program costs. This is the

fundamental policy dilemma facing those seeking to change tax and transfer policies. Before undertaking this effort, it is important to understand the exact nature of the labor supply changes induced by changing tax and transfer policy.

2.2. UK tax and welfare programs

There are effectively four important components of the British direct tax and welfare system as it affects labor supply. The first is the individual tax allowance on earned income, below which no direct taxes are paid. Couples in the UK are taxed independently and the tax allowance is also individually based. In 1996, it was £3650 per year (almost $6000) and was sufficiently large to exempt from direct taxes many part-time low wage workers, especially married women. Approximately 36% of working women married to employed men had earnings below this limit. The majority of workers with earned income above this limit pay direct taxes at a flat basic rate, which has fallen from around 33% to 24% in the 15 years to 1996.[1]

The second component is the National Insurance system which acts like a tax on earnings between a lower and an upper limit. This is also individually based, adds between 2 and 9 percentage points to the basic tax rate and is paid in full once earnings rise above the lower limit. Therefore, unlike the basic tax rate, the NI premium is payable on all earnings. Moreover, as NI payments stop at approximately the level of the higher tax rate, the overall tax rate through the direct tax system rarely exceeds 40%. Third is the "in-work" benefit Family Credit described in Section 2.2.1 (reformed and renamed Working Families Tax Credit in the 1998 Budget). The last of the four components is the multitude of largely means-tested income assistance programs that cover unemployment insurance and housing benefits; child support is a flat-rate non-means tested benefit examined in more detail below. Although the welfare system is designed to acknowledge interdependencies in benefit reduction rates so that no effective tax rate exceeds 100%, combining the tax system with the welfare system implies some severe disincentives for work, especially for low-wage families. This motivated the introduction of an "in-work" tax credit.

It is also worth noting that, over this period, the rate of Value Added Tax, paid on all goods except food and children's clothing has risen from 9% in 1979 to 17.5%.[2]

2.2.1. An hours-based "in-work" benefit

An important component of the British tax and welfare system is the "in-work" benefit program called Family Credit (FC). Introduced in 1988 as an extension to Family Income Supplement, it has many features in common with the EITC program in the US. However, elegibility is based on a minimum weekly working hours requirement. The new Working Families Tax Credit, which replaces FC in October 1999, has exactly the same minimum weekly hours requirement. In this respect, the British in-work benefit system has simila-

[1] In 1993, a lower band of 20% was introduced on a relatively small initial part of taxable income. Less than 15% of workers pay tax at the higher 40% rate.

[2] Domestic energy was exempt from VAT but now attracts a reduced rate of 8%.

rities to the Canadian Self-Sufficiency Program (see Card and Robins, 1996). However, it should be pointed out that the SSP, which is only currently operating on an experimental basis, is time-limited and only available to parents with long durations of welfare receipt and unemployment. The FC system was designed to encourage part-time work and to support the income of part-time working parents. It has subsequently been extended with a small supplement for full-time work.

The basic FC scheme is generous but has a high withdrawal or benefit reduction rate. Family Credit becomes payable to individuals in families with children if their working hours exceed 16 per week and their overall income falls below some specified level, which varies with the number and age of children. The credit is then progressively withdrawn at a reduction rate of 70% as income rises (to be lowered to 55% in the WFTC reform). This rate is considerably higher than that for the EITC in the US.

Since the introduction of FC in 1988, the structure of the in-work benefit system has witnessed three major reforms: a reduction in the hours rule from 24 to 16 in 1992, the introduction of childcare disregards to help recipients with child-care costs in 1994, and the introduction of an additional credit at 30 h. During this period, the number of recipients doubled to well over 500,000. The Working Families Tax Credit reform only marginally changes the structure with a more generous level of payment and a lower benefit reduction rate of 55%. Consequently, more individuals in work who would not have received FC will now receive WFTC.

For most low-income individuals, working less than 16 h per week, the income support and housing benefit system renders the budget constraint virtually flat, so that FC can act as an important jump in the in-work income for low-wage working parents. The high benefit reduction rate, however, implies a reasonably flat constraint above 16 h, providing a potentially strong incentive for those working more to reduce their hours. Consequently, questions similar to those of the EITC arise as to the effectiveness of the system.

Since 1980 there has been no earnings-related unemployment insurance in the UK. Benefits for the unemployed, called job-seekers' allowance (JSA), are flat-rate at a level similar to the level of basic Income Support. This is worth about 20% of median full-time male net weekly earnings and is withdrawn at a rate of 100% against earnings provided weekly hours of work are fewer than 16. At higher hours, no income support is available. However, a child benefit of approximately £10 per week per child is payable to all families regardless of income. Consequently, for childless workers with low housing costs, income out of work is relatively low. For families with children, in particular for lone parents, this is not the case.

Fig. 4 shows the implied net government transfers for a single parent earning £4/h with two pre-school children in the UK. The FC at 16 h produces a large jump in net income. The additional supplement at 30 h is also evident. Fig. 5 displays the same budget constraint but in terms of weekly hours of work. This highlights the minimum hours requirement in the British in-work credit system. We assume a rent level of £50 per week. Housing Benefit (HB) is paid to all individuals with a sufficiently low income, and covers all rent whether the individual is in private or public rental housing. Once

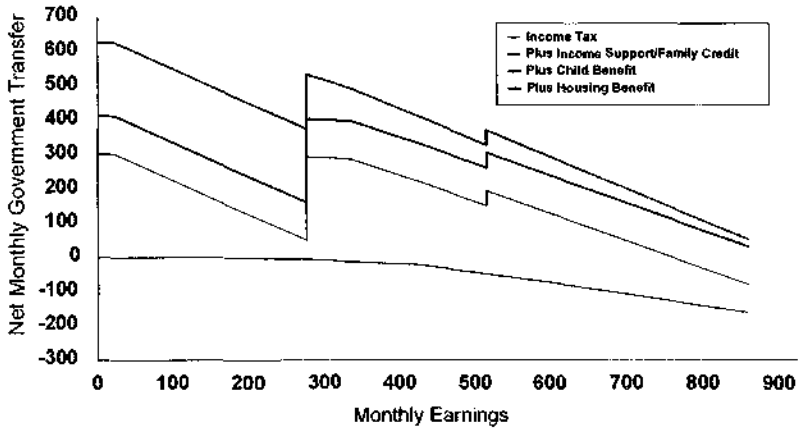

Fig. 4. UK net transfers 1996–1997: lone parent.

income reaches a ceiling, the benefit is withdrawn at a rate of 65%. This is further enhanced since the 65% withdrawal is made after income tax, NI and local taxes have been paid. Income support and other benefits, such as one-parent benefits, can be seen to fall in line with the increase in earned income up to 16 h per week. After that point, FC enters. National Insurance payments also become important and the budget constraint is further flattened by the high benefit reduction rate for FC. The total disposable income line in Fig. 5 shows the combined impact of the UK tax and benefit system.

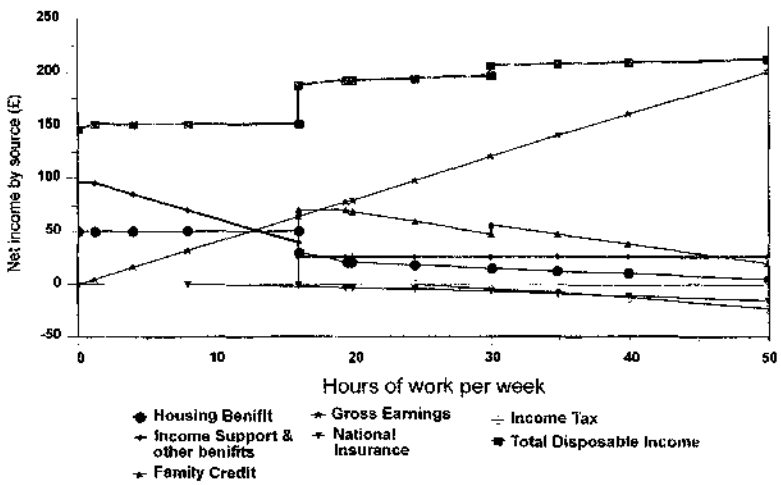

Fig. 5. UK budget constraint, 1996–1997

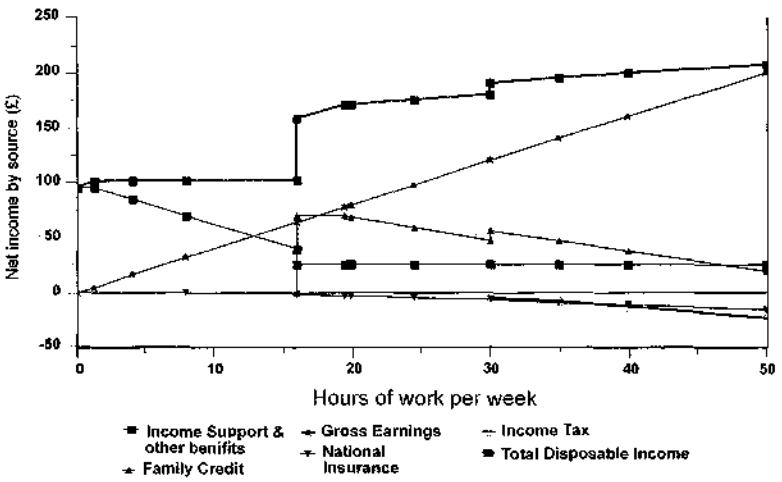

Fig. 6. The impact of removing housing benefit.

2.2.2. Interactions among British programs: family credit, housing benefit and income support

As we noted above for the US, for the purposes of analyzing labor supply it is important to recognize the interactions between benefits and in-work credits. These are critically important for low-wage families and raise similar issues to those discussed in the context of the EITC program. In Fig. 6, we show the impact of excluding Housing Benefit (HB). Since FC is treated as income in determining eligibility for HB, the impact of FC is considerably reduced by the HB program. A comparison of Figs. 5 and 6 shows there is now a much larger increase in net income at 16 h.

Since 1986, Family Credit, Housing Benefit and Income Support have all become important components of the British welfare system. This is revealed in a comparison with Fig. 7 which shows the net government transfers for the same lone parent facing the 1986–1987 welfare and tax system. Lower housing benefits (a mean of £35 rather than £50 per week in 1996 prices) in 1986 reflect the lower level of social rents in public sector housing which, paradoxically, reduced the incentive problem facing low-wage workers. It is probably the rise in public housing rents together with the decline in the relative real wages of low-skilled workers that most significantly changed the balance between work and unemployment for low-wage families in the UK.

3. Recent empirical trends

In conjunction with the large changes in tax and benefit policies detailed in Section 2, the 1980s and early 1990s have seen dramatic changes in participation, hours of work and

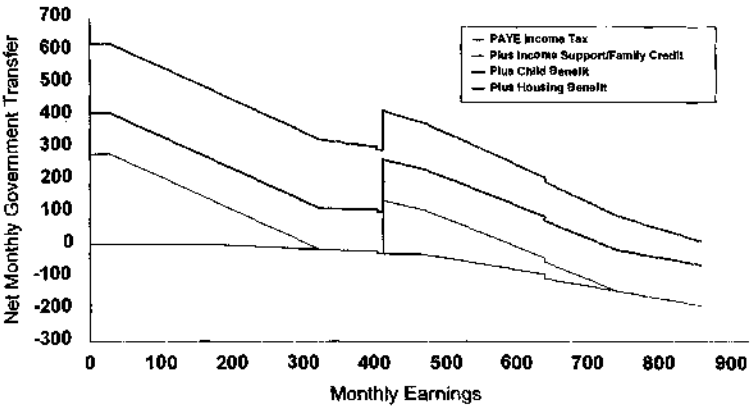

Fig. 7. UK net transfers 1986–1987: lone parent.

hourly wages. In this section, we provide a brief documentary of these changes, drawing on evidence from the US, the UK, Germany and Sweden.

It is the changes in participation and working hours that labor supply models attempt to explain. The success of these models must therefore be judged according to their ability to explain and enhance our understanding of the changes in participation and hours. Moreover, movements in the structure of real wages, in addition to reforms of the tax and benefit system, provide the variation needed to explain these changes. To the extent that trend differences in real wages, government transfers and marginal tax rates across groups can be argued to be exogenous to changes in preferences for labor supply, they provide the most convincing data, outside social experiments, for recovering reliable estimates of labor supply responses. This explains the central role we place on these empirical regularities in this survey.

The changes in participation, hours and real wages have varied widely across economic and demographic groups. For example, higher-educated workers in the UK and US have seen strong growth in real wages, while less-educated workers have experienced stagnant or falling real wages. In contrast, the real wages in all education groups in Germany appear to have risen steadily during this period. There have also been strong differences in labor market attachment across age groups. An increase in the overall participation of women has been matched by a drop off in the participation of males, particularly pronounced among older men in Europe.

In this section, we first discuss changes in participation. These are analyzed by education level for men and women separately. The contrast by education group is striking, as are the differences in the trend changes between men and women. Next, we move to an analysis of hours of work to highlight the changes in the average weekly and annual hours worked by different education and gender groups since the end of the 1970s. Finally, we consider changes in gross hourly wage rates. The detailed changes in these are documen-

ted elsewhere in this Handbook (see the chapter by Katz). However, our aim is to focus on contrasts by education and gender and to evaluate differences in these between the US, the UK, Germany and Sweden (Figs. 8–14).

3.1. Data sources

We draw from a variety of country-specific data sources. Our samples contain men and

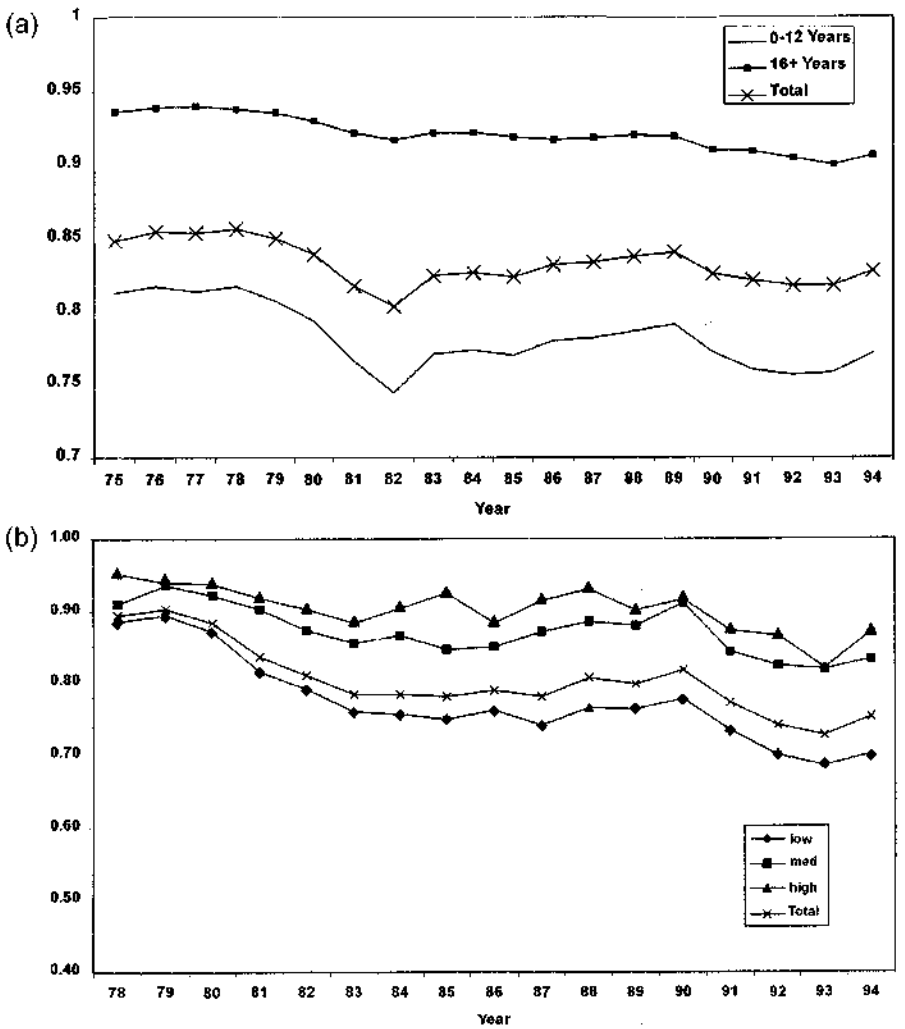

Fig. 8. Men's employment to population ratio by education: (a) US; (b) UK; (c) Germany; (d) Sweden.

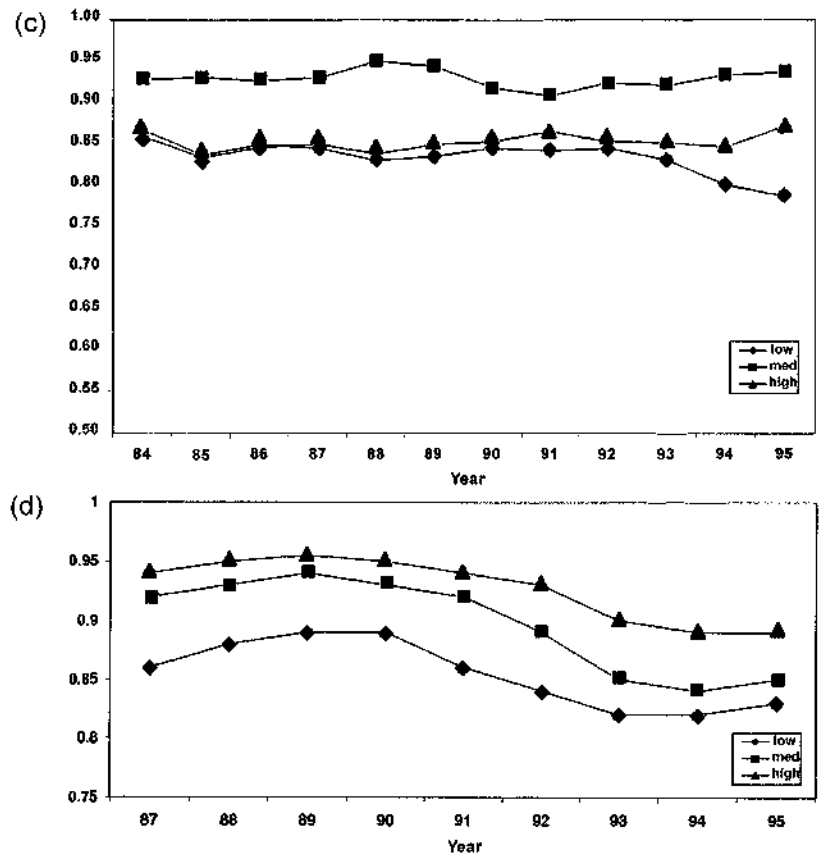

Fig. 8 *(continued).*

women above minimum school leaving age and below the standard retirement age. Participation is defined as the proportion in employment out of all individuals of working age in a specific group. For the US, the primary source of data is the Current Population Survey, a monthly survey of approximately 60,000 households. A group of CPS interviewees stays in the sample for 4 months, is out of the sample for the next 8 months, and then returns to the sample for the following 4 months. We consider data from 1975 to 1994, for men aged 26–64 and for women in multiple-year birth cohorts ranging from 1920–1926 to 1950–1964. For the UK it is the Family Expenditure Survey (FES), a repeated cross section. Each FES survey consists of around 7000 households. All individuals aged between 18 and 59 years of age are used except those in full-time education, self-employment or the armed services. The "low" education group includes those that left formal schooling at the minimum school leaving age (currently age 16). The "med" education group includes those in schooling until age 18. The "high" group includes those with college education.

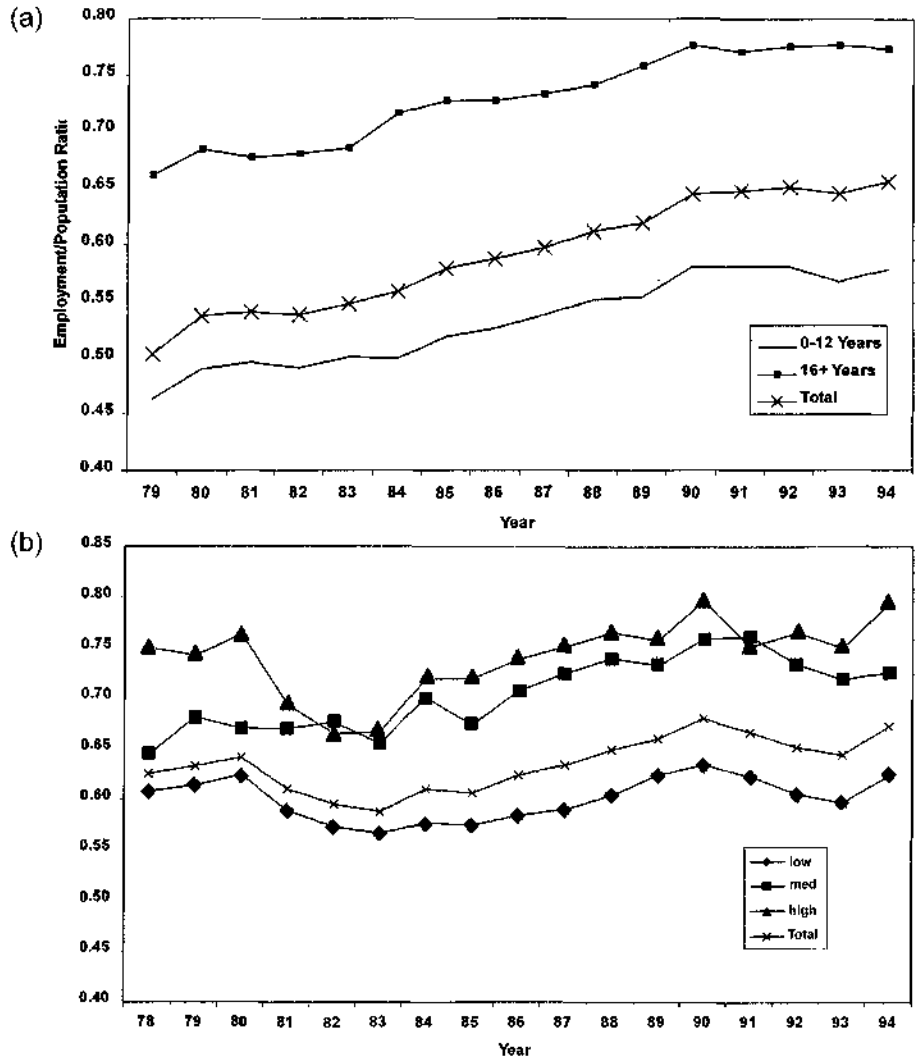

Fig. 9. Women's employment to population ratio by education: (a) US; (b) UK; (c) Germany; (d) Sweden.

For Germany, a similar selection criterion is used and we draw on individuals from the first 12 waves (1984–1995) of the German Socio-Economic Panel (GSOEP). The design of the GSOEP is similar to that of the US Panel Study of Income Dynamics (PSID), see Wagner et al. (1993). All figures reported below refer to individuals located in the geographic area of the former West Germany. The precise details of the data construc-

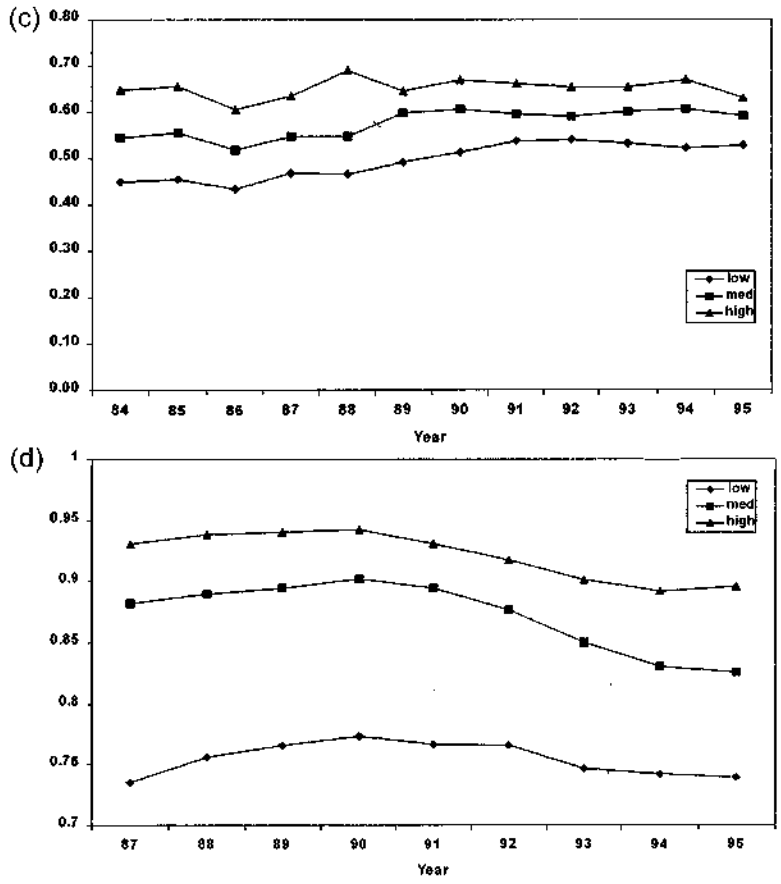

Fig. 9 *(continued).*

tion follow the work of Dustmann and Van Soest (1997). For Sweden, three different data sources have been used. An income survey (HINK) and the Swedish Labor Force Survey (AKU), both from Statistics Sweden. This is supplemented with data from the Swedish survey, Market and Non-Market Activities (HUS) (see Flood et al., 1997 for details).

3.2. Participation

Participation in work has seen some important changes since the late 1970s. Fig. 8a provides the evidence for US men by education level (here measured by years of schooling). The cyclical nature of participation for the lower education group and the much lower participation rates stand out clearly in the data. If anything, there is a slight downward

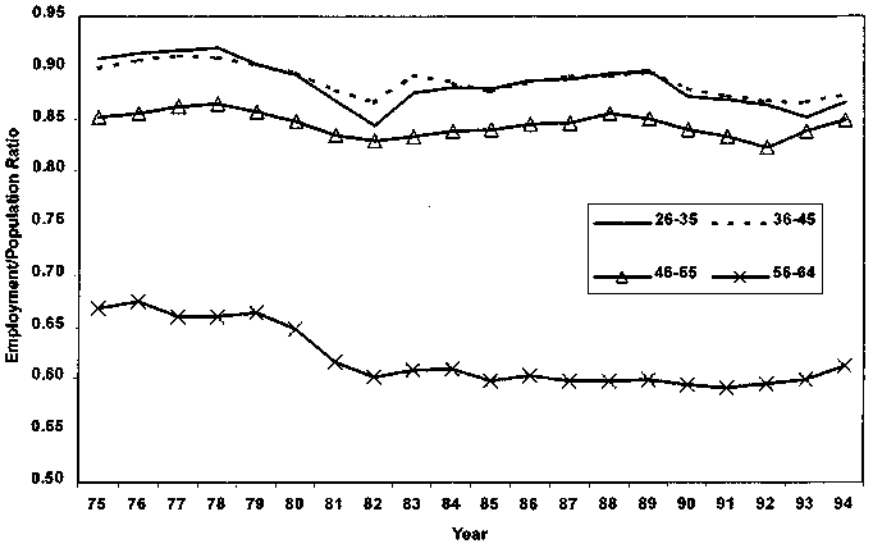

Fig. 10. Men's employment to population ratio by age: US.

trend in participation using this employment-to-population ratio definition. Notice how this differs for women, where Fig. 9a shows that both education groups saw a strong increase in participation until the early 1990s.

This picture for male and female employment in the US has many features in common with the experience in the UK, although, as Fig. 8b shows, male participation has fallen dramatically in the UK since the end of the 1970s. Notice that, even at the top of the boom in 1990, participation did not return to its 1979 levels. This is in contrast to the pattern for women, where Fig. 9b reveals that the participation rate approached 70% in the 1990 boom. In Germany, participation throughout the late 1980s was much more stable than it was in either the US or the UK. Fig. 8c shows that the fall in participation among lower-educated men in Germany only set in after 1992. Remember that these data refer to the West German region both before and after reunification. For German women (Fig. 9c), participation has been slowly rising for all groups until 1992. Finally, in Sweden, we only have a consistent split by education available on an annual basis after 1987. However, until that point, participation rates rose steadily for women and stayed fairly flat for men. The onset of the 1991 recession in Sweden is clear from Figs. 8d and 9d.

The decline in participation for men, which has been experienced to some degree in all countries, is particularly reflected in the working behavior of older age groups. For example, Fig. 10 shows a strong fall in the US employment-to-population ratio for men in the 56–64 year old age group. This declining attachment to the labor market by older men is mirrored in the UK and Germany (see Blundell and Johnson, 1998; Borsh-Supan and Schnabel, 1998). Interestingly, for the UK and Germany, it is the younger birth cohorts

as they age that are seeing larger declines. For older women, this picture is attenuated by the steady rise in participation across time and across birth cohorts. In the US and the UK, there has been an increase in participation for younger birth cohorts of women and, consequently, at the same age, younger cohorts of women have higher participation rates.

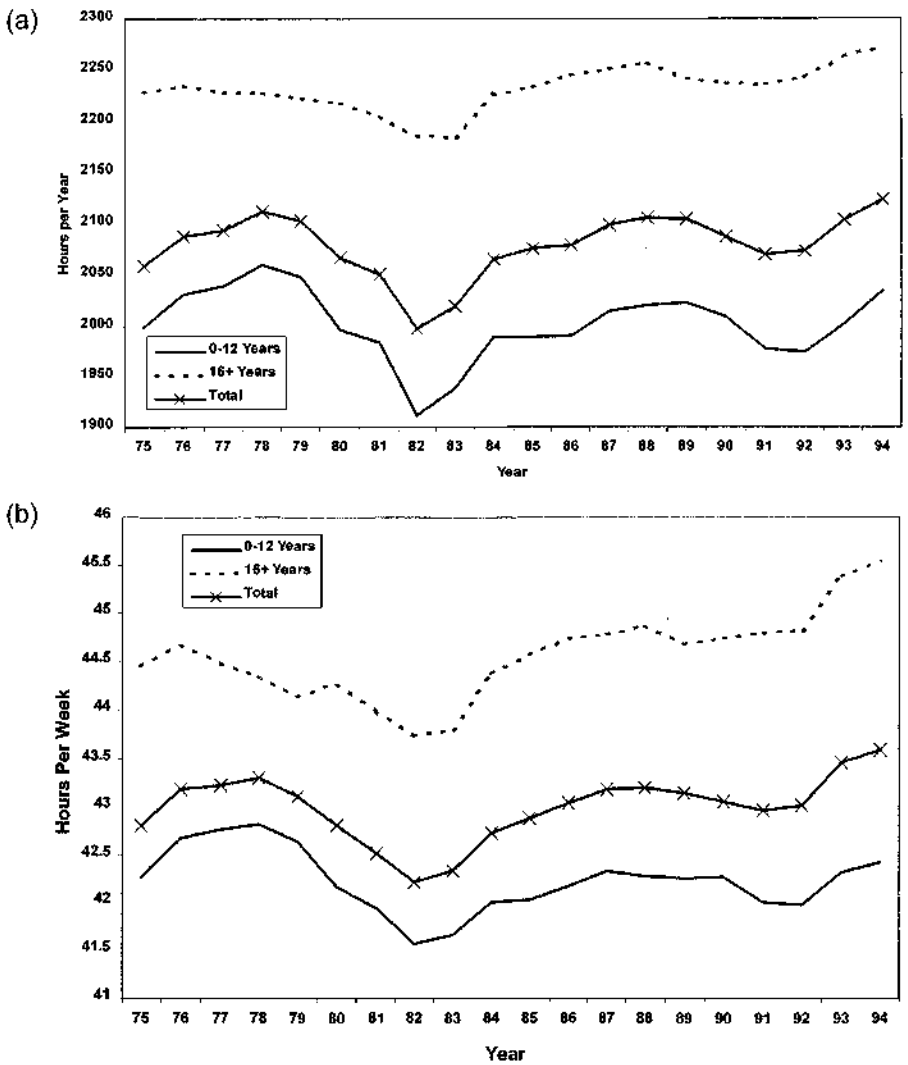

Fig. 11. (a) Men's annual hours worked by education level: US. (b) Men's weekly hours worked by education level: US. (c) Men's weekly hours worked by education level: UK. (d) Men's weekly hours worked by education level: Germany.

3.3. Hours of work

Annual hours of work for men in the US display a strong cyclical pattern and, especially during the last decade, an increasing trend. A similar story is true for weekly hours. These two measures of working hours in the US are presented in Fig. 11a,b. For the UK, the nature of our survey data means that we can only present weekly hours (that include

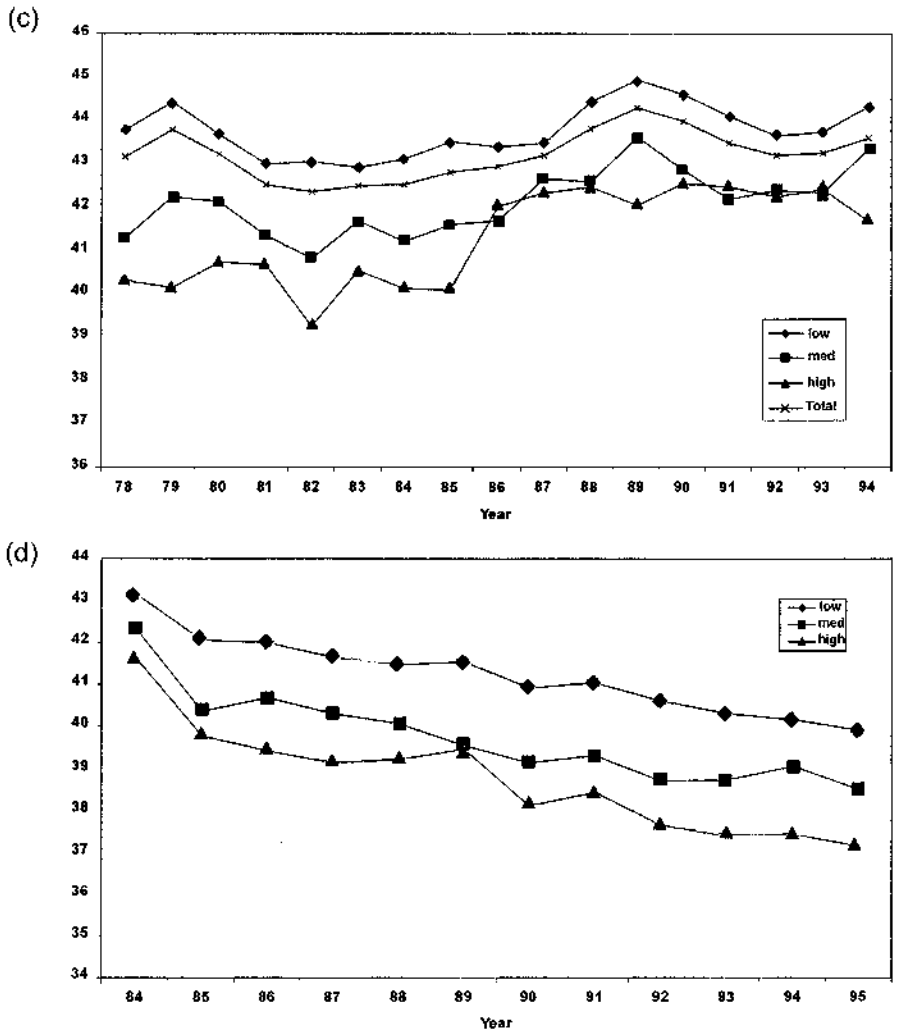

Fig. 11 *(continued).*

normal overtime hours). Fig. 11c shows that this measure of hours worked reveals a similar strong cycle and trend increase although, in contrast to the US, it is the higher-educated group in the UK that has tended to work fewer weekly hours on average. What is notable in both of these countries is that the trend increase in weekly hours is more accentuated for the higher-education group. Interestingly, as we shall see below, this is precisely the group that has seen a trend rise in real wages.

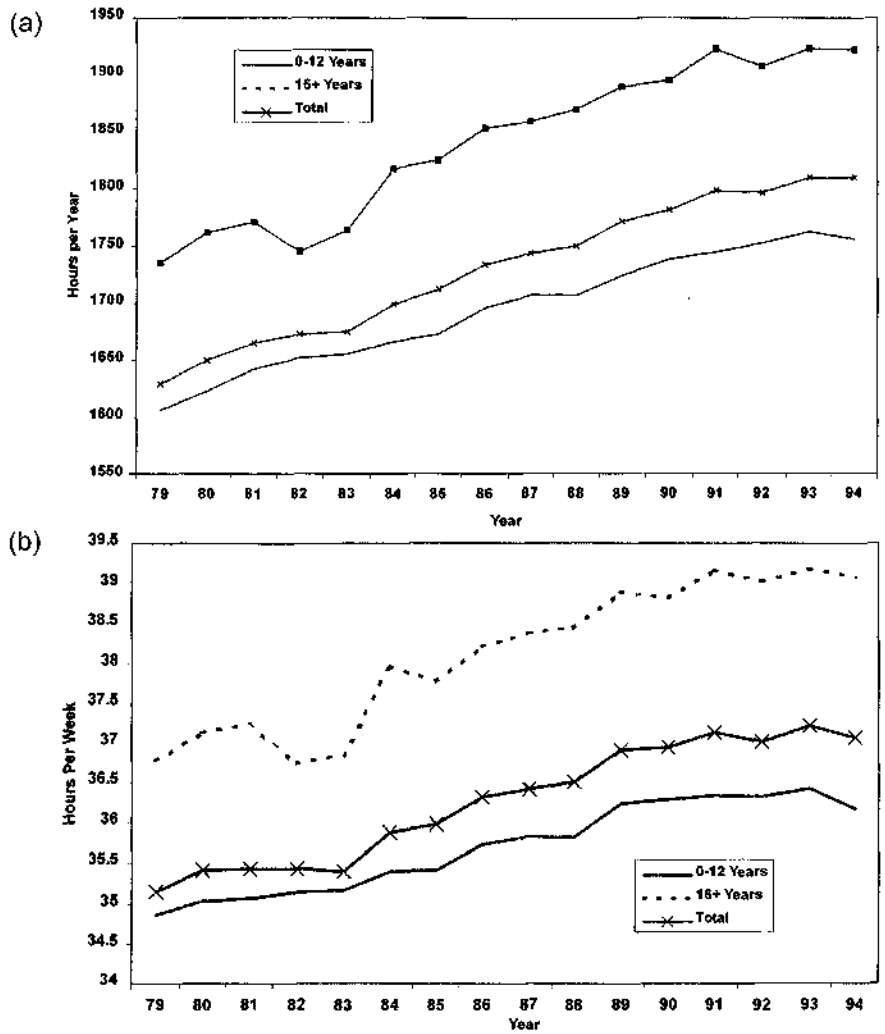

Fig. 12. (a) Women's annual hours worked by education level: US. (b) Women's weekly hours worked by education level: US. (c) Women's weekly hours worked by education level: UK. (d) Women's weekly hours worked by education level: Germany.

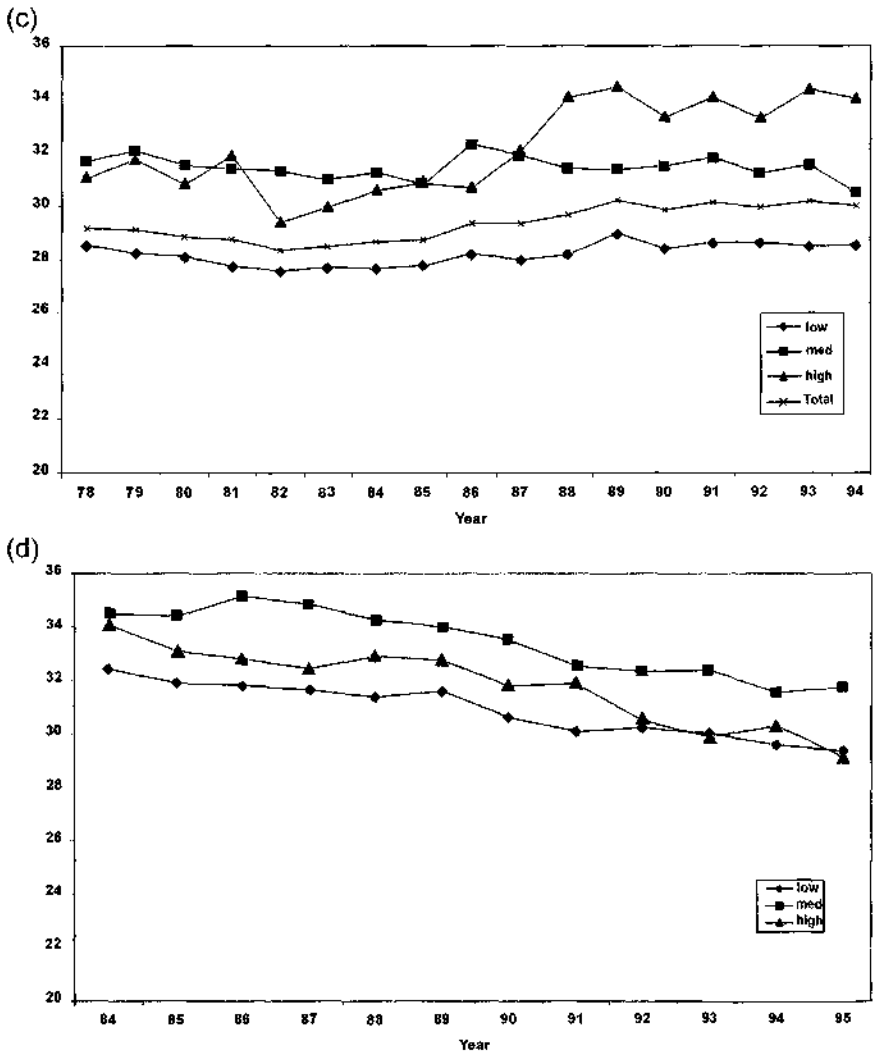

Fig. 12 *(continued).*

Fig. 12a–c shows that a similar story is true for the weekly working hours of women in both the US and the UK, although it is the higher-educated group that works longer weekly hours in the UK. Fig. 12b shows that, if anything, this gap has grown during the recent past. Annual hours have shown a strong trend increase in the US, as seen in Fig. 12a. In the UK this is probably less pronounced, at least for weekly hours of work. None the less, the

UK has seen a steady rise in women's weekly hours since the early 1980s when the cyclical downturn in 1980 and 1981 had a depressing effect on female and male hours of work alike. Although not reproted here, working hours in Sweden for employed males have been quite stable despite a major tax reform in 1991. After 1993 there is a small increase for highly educated workers. For females there has been an upward trend in hours. This is especially pronounced for the highly educated. Working hours in Germany have seen a slow and smooth decline, as evidenced in Figs. 11d and 12d.

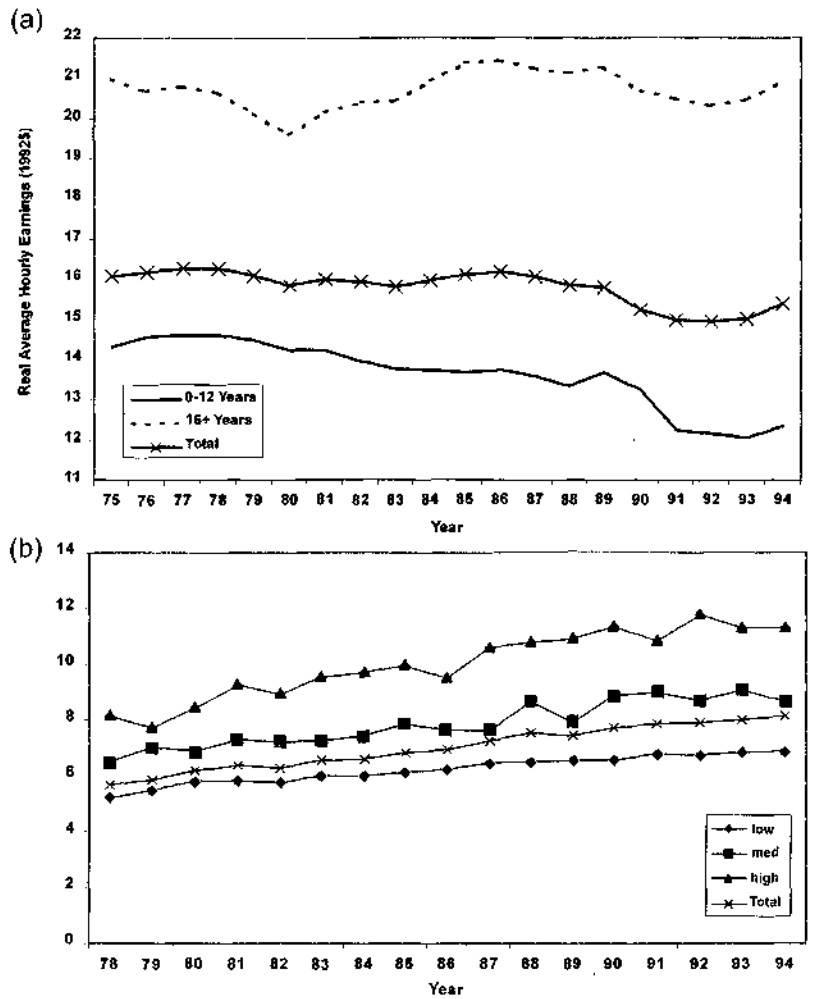

Fig. 13. (a) Men's real average hourly earnings by education level: US. (b) Men's real average hourly earnings by education level: UK. (c) Men's real average hourly earnings by education level: Germany.

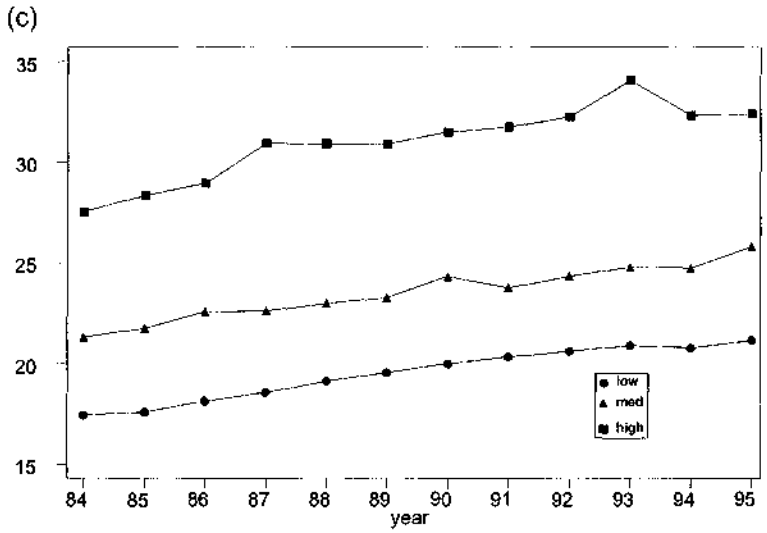

Fig. 13. (*continued*)

3.4. Real wages

The contrasts among the US, the UK, and continental Europe are probably most stark when it comes to a comparison of the level and growth of real wages. This is especially the case when split by education level. However, there are serious pitfalls in the interpretation of raw wage trends. First, there is a considerable change in composition across time both in terms of the total group of employees and in terms of the different education groups, and these composition changes are very different across countries. Second, there is the dubious comparability of definitions of education levels across countries.

The first issue is really at the heart of labor supply analysis itself, since it relates to the changing composition of those in work over time. For example, if lower real wages at the bottom of a cycle mean fewer lower-ability workers supplying labor at that time, then this systematically biases upwards the real wage at the bottom of the cycle. Similarly, if the increasing levels of non-participation by older men reflect a higher proportion of lower-ability workers leaving employment due to a relatively generous social security and benefit system, then this results in an upward bias in measured real wages and in measured returns to experience for low educated workers. This biases upward the trend increase in real wages for lower-educated workers and biases downward the apparent return to education. Any comparison of the growth of real wages and returns to education between the US and European economies must therefore acknowledge the impact of differential changes in composition on real wages.

With these points in mind, turning first to the US, Fig. 13a for men tells a dramatic story. For the lower-education group, real wages have fallen almost relentlessly since the late 1970s. Consequently, the education differential has widened significantly. As Fig. 14a shows, this is less clear-cut for women but, given the rise in participation for the lower-education group of women, a comparison over time may be less interpretable. None the

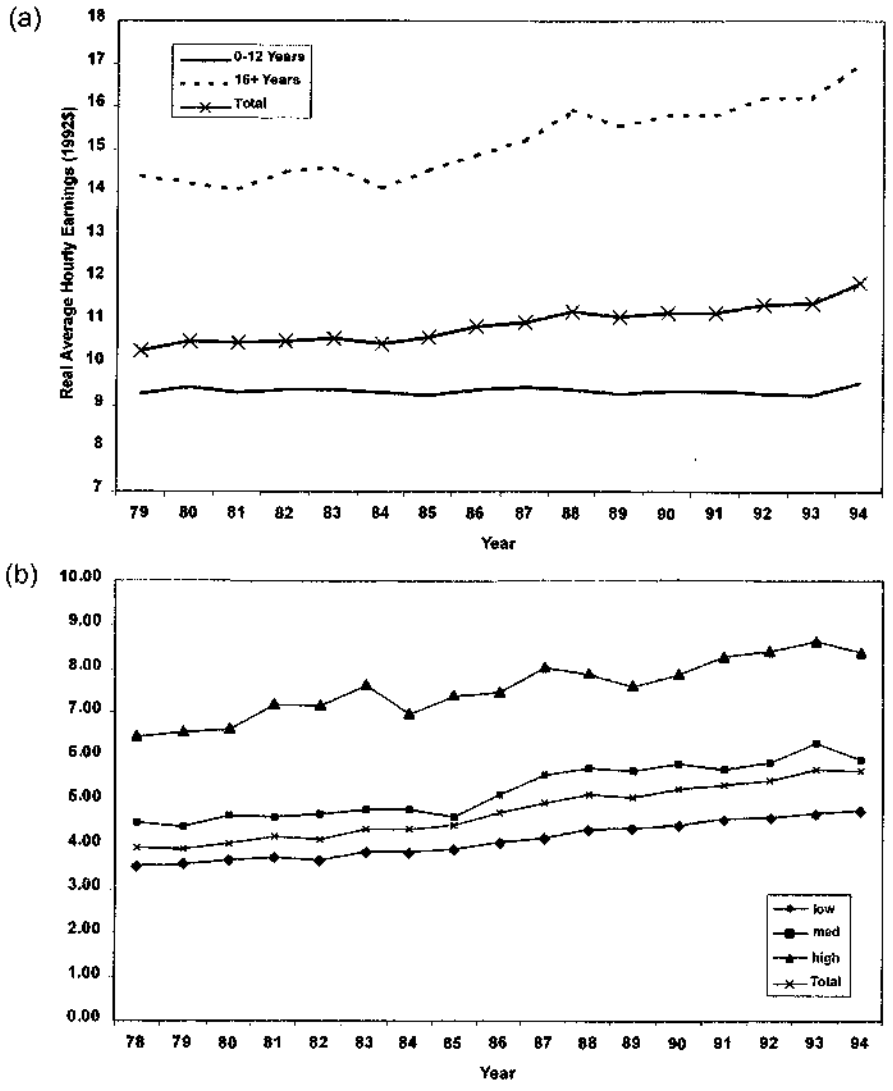

Fig. 14. (a) Women's real average hourly earnings by education level: US. (b) Women's real average hourly earnings by education level: UK. (c) Women's real average hourly earnings by education level: Germany.

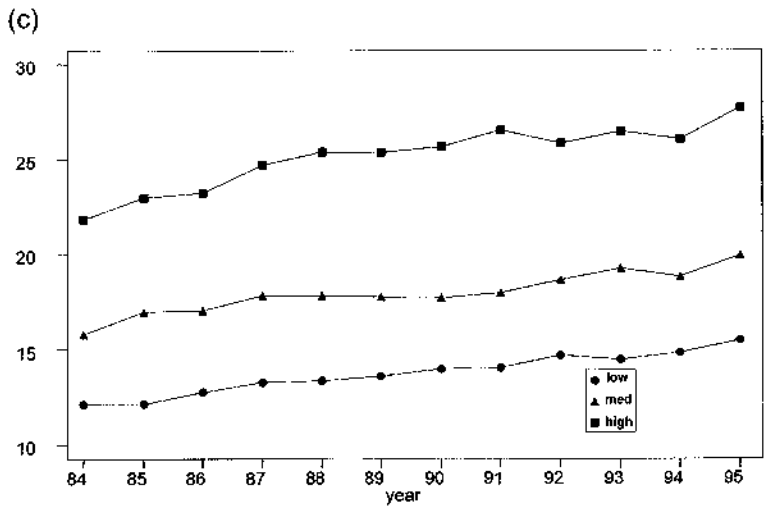

Fig. 14. (*continued*)

less, the increase in the differential is clear and the rise in the real earnings of the higher-educated women is quite spectacular, with a consequent fall in the raw gender differentials. For the UK, Fig. 13b shows an increase in the educational differential for men but, in contrast to the US, no fall in real wages for the lower education group. Although this lower-education group refers only to individuals who left school at 16 or earlier, it still makes up nearly 70% of the UK sample. For this group, composition changes are likely to be quite severe since, as we have already seen, there was a dramatic fall in participation during this period. By contrast, in Sweden there has been an increase in wages for low-educated workers. The wage difference has decreased over time. In Germany, we also see no decline in real wages for the lower-education group. Indeed, at least for the decade 1986–1995, if anything Fig. 13c points to a slight fall in the raw educational differential for men. Given the stable employment rates during most of this period, it is difficult to attribute this rise to a composition effect.

4. A framework for understanding labor supply

Evaluating and interpreting labor supply estimates requires economic models to provide a context for comparison. Estimates often diverge simply because studies focus on evaluating behavioral responses corresponding to different wage and income effects. Sometimes empirical analyses are not precise about what model underlies their estimates. Is a static or

a lifecycle model used? What do substitution effects hold constant? Does the analysis recognize taxes and joint decision-making by family members? Does the model assume perfect certainty or can it allow for uncertainty? Does it assume a representative agent or is individual heterogeneity allowed? When researchers discover divergence in their labor supply estimates, they frequently cite sampling or data differences to explain discrepancies. Equally important, and often more informative for economists seeking to reconcile them, are the differences in economic frameworks used in the studies.

Many empirical studies of labor supply leave the reader to deduce the underlying model from the set of outcome and control variables incorporated in the analysis. Apart from hourly wages and other income, are controls for lifetime wages included? Is a measure of property income included and, if so, how is it measured? Do researchers account for expected changes in income sources? What demographic characteristics are included as controls? Differences in the included conditioning variables implicitly determine the economic framework as well as the response parameters estimated within that framework. Hence, a clear understanding of the implication of these decisions is necessary for any comparison of divergent estimates.

This section presents a unifying framework in which different basic labor supply models can be compared. By considering existing empirical work in one consistent framework, we can determine whether individual studies estimate meaningful parameters and, if so, which parameters are comparable across papers. Empirical studies will have to contend with practical issues concerning non-linear taxation, measurement error and the discreteness in choices. This section abstracts from these complexities so as to focus on the differences in interpretation across models. These complexities are then taken up in the remaining sections of the paper where specific empirical studies are also reviewed.

The development of a unifying structure for interpreting labor supply studies should not suggest that there is one correct way to estimate labor supply equations. Quite to the contrary, we recognize that many of the differences across existing empirical models reflect differences in data availability; our approach seeks to provide a synthesis in which results from each data source can be compared. Data vary in the forms of income that are included, the definition of hours and wage variables, and whether observations are longitudinal or cross-sectional, but meaningful and comparable results can be derived from each if the implications of the estimated function are carefully considered. An understanding of labor supply is greatly enhanced by any available source of exogenous wage and income variation, and no dataset providing this information should be discarded. At the same time, results from varying studies must be comparable, and the framework presented here seeks to facilitate these comparisons.

4.1. The static labor supply model

To set the scene, we begin by outlining the standard static, within-period labor supply model. This is an application of basic consumer theory. Assume each individual has a quasi-concave utility function

$$U(C_t, L_t, X_t) \tag{4.1}$$

in which C_t, L_t, and X_t are within-period consumption, leisure hours and individual attributes, in period t.[3] Utility is assumed to be maximized subject to the budget constraint

$$C_t + W_t L_t = Y_t + W_t T, \tag{4.2}$$

where W_t is the hourly wage rate, Y_t is non-labor income, T is the total time available and a single consumption good is taken as the numeraire. The right-hand side of (4.2), then, includes the full value of one's endowment of time as well as all other sources of income. This is often defined as "full income" from which the consumer purchases consumption goods and leisure. We denote this income concept as M_t, so that

$$M_t \equiv Y_t + W_t T. \tag{4.3}$$

In static models, non-labor income, Y_t, is typically the sum of two components: asset income and other unearned income. We return to the measurement of non-labor income in our analysis of multiperiod models below.

First-order conditions take the familiar form

$$U_C(C_t, L_t, X_t) = \lambda_t, \qquad U_L(C_t, L_t, X_t) \geq \lambda_t W_t, \tag{4.4}$$

where λ_t is the marginal utility of income. If the inequality in (4.4) holds strictly then the individual is not working and $L_t = T$. The wage, W_{Rt}, such that $U_L(Y_t, T, X_t) = \lambda_t W_{Rt}$, is the reservation wage below which the individual will not work.

Many have the mistaken impression that labor supply analyses rely on the assumption that individuals can freely choose their hours of work at a fixed wage with a single employer. The behavioral models considered here can readily be thought of as characterizing situations wherein persons choose their hours of work by selecting across employers offering different wage packages. In such instances, the labor supply function approximates the "average" relationship describing consumers' preferences for work hours and hourly earnings. Moreover, one can also allow for "wages" to vary as a function of hour of work with relatively straightforward modifications of the subsequent analyses.

4.1.1. Alternative representations of labor supply

An equivalent expression for the labor supply conditions (4.4) can be given in terms of the marginal rates of substitution (*MRS*). Eliminating λ_t from the first-order conditions (4.4) yields the equation

$$U_L/U_C \equiv MRS_L(C_t, L_t, X_t) \geq W_t. \tag{4.5}$$

This equation contains all information necessary to relate the level of leisure to the level of consumption.

[3] X_t includes all consumer attributes in this specification – observed and unobserved. Since many individual attributes will not be fully observed by the econometrician, it is important to consider the treatment of unobserved heterogeneity in analyzing empirical specifications.

Solving the first-order conditions yields the Marshallian demand functions:

$$C_t = C(W_t, M_t, X_t), \qquad L_t = L(W_t, M_t, X_t) \leq T. \tag{4.6}$$

Equivalently, using $H_t = T - L_t$ and the definition of M_t in terms of Y_t, we have the hours of work rule,

$$H_t = H(W_t, Y_t, X_t). \tag{4.7}$$

Many empirical studies of labor supply seek to estimate forms of (4.7). They vary widely in the measurement of the wage W_t, the income variable Y_t and the demographic controls incorporated in the specification. Depending how these issues are resolved, our full lifecycle framework, developed below, shows that "static" estimates can represent several types of substitution and income effects, ranging from those predicting responses to intertemporal movements in wages to those predicting responses to shifts in entire wage profiles.

Studies generally focus on the wage elasticity of the Marshallian supply function in (4.7), and on the associated utility-constant Hicksian wage elasticity. The Marshallian (uncompensated) wage elasticity is defined as

$$K_u = \partial \ln(H_t)/\partial \ln(W_t). \tag{4.8}$$

Denoting the Hicksian (compensated) wage elasticity by K_c, the Marshallian and Hicksian wage elasticities are linked by the Slutsky equation

$$K_u = K_c + \frac{W_t H_t}{Y_t} \frac{\partial \ln(H_t)}{\partial \ln(Y_t)}, \tag{4.9}$$

where the share $W_t H_t / Y_t$ is the size of earnings relative to non-labor income. The standard sign, homogeneity, and symmetry restrictions from consumer demand theory apply to the Hicksian supply function and have been used to check on the theoretical predictions of the model. Assuming that leisure is a normal good, this expression implies that the Hicksian compensated elasticity is larger than the Marshallian elasticity – the well known result that income and substitution effects work in opposite directions in Marshallian demand.

4.1.2. Family labor supply

Placing labor supply in a family or household context adds a number of important dimensions. Many tax and benefit policies designed to influence labor supply behavior can only be properly understood within a family labor supply framework. Moreover, the changes in the structure of wages facing men and women presented in Section 3, as well as changes in fertility, have important consequences in understanding the changing balance between men and women in family labor supply.

The standard "unitary" family labor supply model treats the family as a single decision-making unit. The attractions of this formulation are that standard welfare results from consumer theory are available and that the family labor supply model can be placed easily within the intertemporal framework. However, although with sufficient separability in

household members' utility the unitary approach can allow for decentralization of within-household allocations, these allocations continue to satisfy the Slutsky symmetry restrictions from consumer theory and also the "income-pooling" restrictions in which the marginal value of non-labor income is equalized across decision-making units within the family. Slutsky symmetry and income pooling are often considered to be unreasonable restrictions and a popular alternative framework that relaxes these latter two restrictions is the collective family labor supply model. This alternative representation of joint labor supply decisions also implies testable restrictions and is sensitive to the introduction of household production. A full discussion of the collective model and its relationship to the standard unitary framework is presented in Section 7. That section also provides a detailed evaluation of the empirical studies of family labor supply and considers the introduction of non-linear taxation and welfare programs. Here we simply outline the basic family labor supply model.

Suppose a family or household consists of two working-age individuals. Children and any other dependents are included in the vector of household attributes, X_t. Families are assumed to maximize joint utility over consumption, C_t, and the leisure of each family member, L_{1t} and L_{2t}. For such a household, utility may be written

$$U_t(C_t, L_{1t}, L_{2t}, X_t). \tag{4.10}$$

The budget constraint now takes the form

$$C_t + W_{1t}L_{1t} + W_{2t}L_{2t} = Y_t + W_{1t}T + W_{2t}T, \tag{4.11}$$

with full income now given by

$$M_t = Y_t + W_{1t}T + W_{2t}T. \tag{4.12}$$

The unearned income term, Y_t, combines all sources of non-labor income.

For the present discussion we consider consumption measured as a single aggregate, C_t.[4] The first-order condition for consumption (4.4) continues to hold, but now it governs family consumption. For leisure choices, (4.4) is extended to give

$$\begin{aligned} &U_C(C_t, L_{1t}, L_{2t}, X_t) = \lambda_t, \\ &U_{L1}(C_t, L_{1t}, L_{2t}, X_t) \geq \lambda_t W_{1t}, \\ &U_{L2}(C_t, L_{1t}, L_{2t}, X_t) \geq \lambda_t W_{2t}. \end{aligned} \tag{4.13}$$

Reservation wages can be computed for each family member exactly as above. Demand functions now take the form

[4] Typically, consumption, even for privately consumed goods, is recorded at the household level. Consequently, measuring individual consumption is difficult. However, for goods such as clothing, separate measurement is often recorded and can be helpful in identifying individual preferences within a household. This is described further in Section 7.

$$C_t = C(W_{1t}, W_{2t}, M_t, X_t),$$

$$L_{1t} = L_1(W_{1t}, W_{2t}, M_t, X_t) \leq T, \tag{4.14}$$

$$L_{2t} = L_2(W_{1t}, W_{2t}, M_t, X_t) \leq T.$$

This model provides a useful framework for thinking about household labor supply decisions. Clearly, if utility is weakly separable in the individual leisures then, provided the appropriate definition of income is used, individual labor supplies can be modeled in the usual way. However, separability is a strong restriction and one that would typically fail in a model that allowed for household production. In Section 7 we describe the family labor supply model more fully, particularly its relation to collective models of labor supply, household production and the analysis of discrete choices. We also present an overview of recent empirical results on family labor supply.

4.2. Multiperiod models of labor supply under certainty

Although its study is often placed in a static framework, labor supply is clearly part of a lifetime decision-making process. Individuals attend school early in life, accumulate wealth while in the labor force, and make retirement decisions late in life; each of these activities can only be understood in a lifecycle framework. We know that savings from labor earnings are often required to sustain individuals, or their dependents, during periods when they are out of the labor market. In addition, variations in health status, family composition and real wages provide incentives for individuals to vary the timing of their labor market earnings for income-smoothing and insurance purposes. In this section, we present the basic components of a single-agent lifecycle labor supply model assuming perfect certainty.

A full lifecycle model, starting at time t, is characterized by a utility function of the form

$$U_t = U(C_t, L_t, X_t, C_{t+1}, L_{t+1}, X_{t+1}, \ldots, C_\tau, L_\tau, X_\tau). \tag{4.15}$$

The intertemporal budget constraint can be represented by the time path of assets, A, as

$$A_{t+1} = (1 + r_{t+1})(A_t + B_t + W_t H_t - C_t), \tag{4.16}$$

where A_{t+1} is the real value of assets at the beginning of period $t + 1$, r_{t+1} is the real rate of return earned on assets between t and $t + 1$, and B_t represents unearned-non-asset income. Individuals maximize (4.15) subject to the series of constraints given by (4.16) for all t through some fixed horizon τ; τ is assumed to be known for simplicity.

This full model is empirically intractable, so virtually all studies assume some form of separability in time. That is, they assume that the utility function can be written

$$U_t = U(U^t(C_t, L_t, X_t), U^{t+1}(C_{t+1}, L_{t+1}, X_{t+1}), \ldots, U^\tau(C_\tau, L_\tau, X_\tau)). \tag{4.17}$$

In this case, the marginal rate of substitution between leisure and consumption in period s can be written as

$$MRS_{Ls} \equiv (\partial U/\partial L_s)/(\partial U/\partial C_s) = U_L^s/U_C^s. \tag{4.18}$$

Combining this with the intertemporal budget constraint, we see that a necessary condition for maximization is

$$MRS_{Ls} \gtrless W_s. \tag{4.19}$$

So, with separability across time, the within-period marginal rate of substitution condition continues to characterize the relative amounts of leisure and consumption. All that remains, then, is to find a summary statistic that captures the impact of other periods on this decision, thus allowing one to pin down the levels of leisure and consumption. The two common methods for this are two-stage budgeting and marginal-utility-of-wealth-constant labor supply.

4.2.1. Two-stage budgeting

The idea behind two-stage budgeting is simple.[5] Since the within-period marginal rate of substitution conditions continue to characterize behavior, we only need an allocation of full income, M_t, to each period to allow each maximization problem to be solved exactly as it was in the static problem. Hence, the decision rule can be decomposed into two stages: first, determine an allocation of wealth across periods; second, within each period, solve the standard static maximization problem. The solution to this problem can be found by reversing the two stages: first, maximize each period's utility, given some M_t; this yields an indirect utility function, $V_t(M_t, W_t)$, for each period. Then, insert the V_t's into U_t and choose the M_t's to maximize this function given current wealth and future wages (expected wages under uncertainty). This solution can be represented by the demand Eqs. (4.6) and (4.7) together with an equation for M_t:

$$M_t = M(A_{t-1}^*, r_t, W_t, Y_t, X_t, Z_t), \tag{4.20}$$

where we have defined A_{t-1}^* to be the end of period $t-1$ assets[6] and Z_t represents future values of W, Y, r and X.

To compare this specification to the static specification introduced in Section 4.1, notice that the two-stage budgeting model automatically corrects the full income measure for the change in assets appropriate in the multiperiod model. From the definition of M_t, we may write

$$M_t = C_t + W_t L_t = r_t A_{t-1}^* + \Delta A_t^* + B_t + W_t T,$$

where now $r_t A_{t-1}^*$ is the real interest income available for expenditure on consumption at

[5] Gorman (1959, 1968) is widely credited with developing the full implications of two-stage budgeting. MaCurdy (1983) and Blundell and Walker (1986), among others, have applied this concept in empirical analyses of labor supply.

[6] In the discussion of multiperiod models, we use two definitions of assets: A_{t+1} in Eq. (4.16) is the beginning of period $t+1$ assets and is therefore equal to $(1 + r_{t+1})A_t^*$. We do this to make our definition of the within-period budget for the two-stage budgeting problem consistent with the intertemporal constraint on assets (4.16).

the beginning of period t and ΔA_t^* is the adjustment in the level of real assets by the end of period t. In contrast, the full income variable in the static model simply includes real interest income and other non-asset income and is given by

$$Y_t + W_t T = r_t A_{t-1}^* + B_t + W_t T,$$

omitting the term ΔA_{t-1}^* which captures the intertemporal adjustment in assets.

The hours-of-work rule in the two-stage budgeting framework, mirroring (4.7), has the form

$$\max U(C_t, H_t, X_t)$$

subject to the budget constraint

$$C_t + W_t H_t = Y_t^C,$$

where we define the consumption-based other income variable

$$Y_t^C \equiv r_t A_{t-1}^* + \Delta A_t^* + B_t.$$

The first stage allocation (4.20) becomes

$$Y_t^C = Y^C(A_{t-1}^*, r_t, W_t, B_t, X_t, Z_t). \tag{4.20'}$$

Note the appeal of the two-stage budgeting formulation. If consumption and leisure (work) hours for the period are observed, then M_t is observable via the within-period budget constraint

$$M_t = C_t + W_t L_t. \tag{4.21}$$

The appropriate adjustment of full income M_t or other income Y_t can be made either with information on assets across periods or with information on consumption. Hence, given some specification for the expectational variables Z_t, one can estimate (4.6) and (4.20) just as in the static framework. Marshallian elasticities can be derived by conditioning on Y_t^C in place of Y_t and can be converted to compensated elasticities via the Slutsky equation, yielding estimates of all other response parameters of interest. Of course, even if the static model were true, or some variant in which there were borrowing restrictions, the within-period allocations that condition on the consumption based measure of full income remain valid.[7]

In evaluating studies using this framework, one must keep two important considerations in mind. First, the appropriate measure of income is the value of consumption plus the value of leisure – that is the full income allocated to the period. Many researchers perform their analyses in a static framework, arguing that they are estimating the second stage of the two-stage budgeting process, but define income as current wages plus unearned

[7] Note that this specification places no restrictions on the path of wages or interest rates, so that employment or capital market constraints can be accounted for, with a wage of 0 indicating no acceptable employment opportunities and an interest rate of ∞ indicating completely constrained capital markets.

income. As we have shown, in a lifecycle setting, these current income figures are irrelevant to current period work and consumption decisions except in so far as they impact the determination of M_t. Second, often elasticities or other response parameters estimated in this basic framework take M_t as fixed and exogenous, just as it is in the static model. Not only does this require the far-fetched notion that consumption is exogenous (if full income is valued with a consumption measure), but it misses any of the response to shocks that occurs through the first stage – that is, through a reallocation of the M_ts. In general, it is only by estimating both stages of the intertemporal allocation model that such responses can be fully accounted for. We take up this issue further in our discussion of multiperiod models under uncertainty.

4.2.2. Frisch labor supply equations and the Euler condition

Marginal-utility-of-wealth-constant labor supply functions, known as Frisch functions, provide an alternative and extremely useful method for analyzing lifecycle maximization problems. In this framework, the marginal-utility-of-wealth parameter, λ_t, serves as the sufficient statistic which captures all information from other periods that is needed to solve the current-period maximization problem. Our discussion critically relies on intertemporal strong separability in preferences, and, for simplicity, the analysis assumes a non-stochastic interest rate.

A useful representation of the problem is given by the functional equation formulation of dynamic programming. Consumers choose consumption and leisure according to the value function

$$V(A_t, t) = \max[U(C_t, L_t, X_t) + \kappa V(A_{t+1}, t+1)] \tag{4.22}$$

subject to the asset accumulation rule (4.16). κ represents the consumer's discount factor. Standard dynamic programming techniques yield the following first-order conditions:

$$U_c(C_t, L_t, X_t) = \lambda_t,$$

$$U_l(C_t, L_t, X_t) \geq \lambda_t W_t, \tag{4.23}$$

$$\lambda_t = \kappa(1 + r_{t+1})\lambda_{t+1},$$

where λ_t is the marginal utility of wealth, $\partial V/\partial A_t$. These are the same first-order conditions as in the static problem, with the addition of the Euler equation for λ. This equation is central to the solution method since it determines the rule for the allocation of wealth across periods. In this formulation, the consumer chooses savings so that the marginal utility of wealth in period t equals the discounted value of the marginal utility of wealth in period $t+1$, where the rate of discount is $\kappa(1 + r_{t+1})$.

These first-order conditions imply consumption demand and hours-of-work supply functions of the form

$$C_t = C(\lambda_t, W_t, X_t), \qquad H_t = H(\lambda_t, W_t, X_t) \geq 0. \tag{4.24}$$

These are commonly referred to as Frisch demand functions.[8] Their functional form depends only on the form of the utility function and whether a corner solution is chosen for hours of work at age t. These functions decompose consumption and labor supply decisions into components observed in the current period, X and W, and λ, which summarizes the relevant information from all other periods. Variables such as future wealth, wages, or personal characteristics affect consumption and labor supply only by changing the value of λ_t. Thus, λ_t serves the role of sufficient statistic, just as M_t was a sufficient statistic in the two-stage budgeting model.[9]

These Frisch labor supply functions are a third type of labor supply function along with the Marshallian and Hicksian functions previously discussed. Whereas Marshallian functions hold income constant and Hicksian functions hold utility constant, Frisch functions hold the marginal utility of wealth constant. One can calculate wage elasticities of the form $\partial H_t/\partial W_t$ for Frisch functions just as for Marshallian and Hicksian functions. We saw above that the Hicksian elasticity is larger than the Marshallian when leisure is a normal good; MaCurdy (1981) and Browning et al. (1985) show that the Frisch elasticity is the largest of the three.

The Euler equation implies a time path for λ of the form

$$\ln \lambda_t = b_t + \ln \lambda_{t-1} \tag{4.25}$$

where $b_t = -\ln(\kappa(1 + r_t))$. Repeated substitution yields

$$\ln\lambda_t = \sum_{j=1}^{t} b_j + \ln\lambda_0. \tag{4.26}$$

Hence, the λ term in (4.24) can be captured as an individual fixed effect, λ_0, plus a function of age which is common across consumers.[10] This ability to model differences in λ as individual effects is very important in the empirical specifications discussed below.

Estimation of (4.24) only allows computation of the Frisch elasticity. This measures the effect of a change in wages holding λ constant. As shown above, in this world of perfect certainty, the path of λ through time is determined solely by the known path of interest rates and the discount factor. Hence, for a given individual, changes in wages have no impact on λ and thus the Frisch elasticity is the correct elasticity for assessing the impact of wage changes through time on labor supply. However, researchers are often interested in comparing the impact of wage variation across consumers on labor supply. In this case, we do not simply examine evolutionary wage changes through time, but rather variation in

[8] We have presented the hours-of-work supply function here, rather than the equivalent leisure demand function. With only two uses for a consumer's time, the two are obviously related by the identity $L_t = T - H_t$.

[9] Note that if labor and capital income is taxed jointly by a non-linear tax, these conditions may need adapting (see, e.g., Blomquist, 1985).

[10] The b_j terms are functions of κ and r which are assumed constant across consumers. Note that if we assume the rate of time preference, ρ (where $\kappa = 1/(1 + \rho)$), equals the rate of interest, b_j is 0 for all j and λ is constant over time.

the entire wage profile. This variation certainly impacts the value of λ_0 and, thus, the Frisch wage elasticity is inappropriate for measuring the effect of such wage variation. To estimate the full impact of wages requires a specification of the impact of the wage profile on λ_0. We consider this further at the end of this section where we evaluate the appropriate elasticity measures for alternative policy questions.

4.2.3. Multiperiod models of family labor supply

The family labor supply model becomes more complicated with the addition of multiple periods or uncertainty, as family composition may change over time. As long as the unitary model is maintained, however, its analysis is straightforward. The marginal conditions for the λ-constant, Marshallian and marginal rate of substitution labor supply equations described above follow naturally from the first-order conditions given by (4.13). Notice that there is still only a single marginal utility of wealth, λ_t, and, therefore, there remains only one Euler condition as in the third equation of (4.23). Consequently, allocations to each individual in this time separable model satisfy equality of marginal utility of wealth. However, to avoid strong separability assumptions between each family member's leisure, careful choice of specification for Frisch labor supplies is required. Further extensions of the multi-period to the family labor supply case are presented in Section 7.

4.3. Multiperiod models of labor supply under uncertainty

The concepts developed in the certainty case essentially carry over to a lifecycle model that recognizes that individuals make labor supply choices in an environment in which they are uncertain about their futures. This requires replacing the deterministic dynamic programming characterization of behavior that we considered in the previous section with a formulation in which agents optimize expected lifetime utility.[11]

4.3.1. Two-stage budgeting under uncertainty

Accounting for uncertainty in two-stage budgeting is inconsequential. Eqs. (4.6) and (4.20) continue to summarize choices. Actually solving for the optimum period-specific expenditure allocation, M_t, is now more difficult since wealth cannot be allocated once at the beginning of life and instead must be reallocated each period as information is revealed. One can solve this problem, however, via standard dynamic programming formulations. Instead of including realized values, the variables Z_t in (4.20) (or (4.20′)) now include attributes of the distribution of future wages and income, and future determinants of preferences. In Section 4.5, we consider approaches to estimation of the full lifecycle model which combine the two-stage budgeting formulation with the intertemporal first-order conditions on consumption. This turns out to be a useful way of characterizing the preference restrictions underlying various empirical specifications.

[11] Much of this framework comes from MaCurdy (1985). We refer the reader to this reference for details of the development of these specifications.

4.3.2. Frisch labor supply under uncertainty

With the introduction of uncertainty over future wages, the dynamic programming representation of the consumer's problem changes to

$$V(A_t, t) = \max\{U(C_t, L_t, X_t) + \kappa E_t[V(A_{t+1}, t+1)]\} \tag{4.27}$$

subject to the asset accumulation rule (4.16).[12] The first-order conditions now include (4.23) and (4.24) together with a modified Euler equation,

$$\lambda_t = \kappa E_t[\lambda_{t+1}(1 + r_{t+1})]. \tag{4.28}$$

The only change from the certainty case is that λ_{t+1} is now a random variable which is not realized until the start of period $t+1$. The savings allocation rule, given by (4.28), determines the path followed by λ through time. Given that the consumer cannot perfectly control the level of his wealth, his environment changes as he acquires information and λ_t is stochastic. Condition (4.28) describes how the consumer allocates his resources to account for unanticipated shocks. He sets his savings policy so that the expectation of next period's marginal utility of wealth is revised by the full amount of the unanticipated elements; in other words, the consumer revises the means of all future values of λ to account for all forecasting errors when they are realized. The standard Euler equation for consumption is derived by replacing λ_t and λ_{t+1} in (4.28) by $U_c(C_t, L_t, X_t)$ and $U_c(C_{t+1}, L_{t+1}, X_{t+1})$, respectively.

A useful characterization of the stochastic process for λ implied by (4.28) takes the form

$$\ln\lambda_t = b_t^* + \ln\lambda_{t-1} + \varepsilon_t^*,$$

where the coefficient b_t^* depends on the discount factor, κ, the interest rate, r_t, and the moments of the forecast error, ε_t^*.[13] Repeated substitution yields

$$\ln\lambda_t = \sum_{j=1}^{t} b_j^* + \ln\lambda_0 + \sum_{j=1}^{t} \varepsilon_j^* = b^* t + \ln\lambda_0 + \sum_{j=1}^{t} \varepsilon_j^*, \tag{4.29}$$

where the last expression assumes $b_t^* = b^*$ for expositional simplicity.

Combining Eq. (4.29) with the consumption and labor supply conditions in (4.23) suggests a simple view of lifecycle behavior under uncertainty. At the start of the lifetime, the consumer sets the initial value of λ_0 to include all available information. As he ages, he

[12] For simplicity of presentation, we continue to assume a non-stochastic interest rate, although the extension to random interest rates is trivial.

[13] To develop this expression, write $\ln\lambda_t = E_{t-1}\{\ln\lambda_t\} + \varepsilon_t^*$. This relation implies $E_{t-1}\{\lambda_t\} = \exp[E_{t-1}\{\ln\lambda_t\}]E_{t-1}\{\exp[\varepsilon_t^*]\}$, which in turn yields $\lambda_t = [E_{t-1}\{\exp[\varepsilon_t^*]\}]^{-1}E_{t-1}\{\lambda_t\}\exp[\varepsilon_t^*]$. Inserting the condition for $E_{t-1}\{\lambda_t\}$ given by Eq. (4.28) into this latter relation yields $\lambda_t = [E_{t-1}\{\exp[\varepsilon_t^*]\}]^{-1} \times(\kappa(1 + r_t))^{-1}\lambda_{t-1}\exp[\varepsilon_t^*]$. Taking natural logs gives $\ln\lambda_t = b_t^* + \ln\lambda_{t-1} + \varepsilon_t^*$ where $b_t^* \equiv -\ln(\kappa(1 + r_t)) - \ln(E_{t-1}\{\exp[\varepsilon_t]\})$.

responds to new information by updating λ according to (4.29). At each age, the consumer only needs the updated λ, along with current wages and characteristics, to determine his optimal consumption and labor supply.

A substantial complication to the labor supply models described in this section arises if we relax the assumption of time-separable utility. For example, consider allowing an individual's wage to be a function of human capital, which this person chooses to acquire by training. In this case, the wage is endogenous, as it is determined by an individual's training decision. The primary method of dealing with this complication is to move to a fully structural model of lifetime decision-making in which parameter values are chosen to match closely the observed work, training and consumption decisions. We defer a discussion of this and other related dynamic generalizations of the labor supply model that relax the time separability assumption to Section 8, which considers dynamic structural models.

4.4. Basic empirical specifications

A prototype empirical specification that encompasses many economic models of labor supply takes the form

$$\ln H_t = \alpha \ln W_t + \beta Q_t + e_t, \tag{4.30}$$

where α and β are parameters, Q_t is a vector of "controls" and e_t is a stochastic term unobservable to the economist. In what follows, we consider alternative specifications for βQ_t.

Studies also often use alternative transformations of H as the dependent variable. For example, a popular alternative is the semi-log specification

$$H_t = a \ln W_t + bQ_t + v_t, \tag{4.30'}$$

which is particularly attractive for dealing with non-participation. One also finds various formulations for wages as right-hand side variables (such as after-tax wages or non-linear functions of wage rates). In each case, additivity between the log wage variable, unearned income variables and the other controls will imply restrictions on preferences. The preference restrictions underlying these and other popular labor supply specifications are reviewed in Appendix A.

The value of α in (4.30) determines the substitution effect associated with the response of labor supply to changes in wages. As discussed above, the interpretation of this substitution effect varies according to precisely which controls one includes in the vector Q_t and which of these controls are treated as exogenous.

4.4.1. Static specifications

The conventional static specification involves estimating Eq. (4.30) with controls set according to

$$\beta Q_t = \rho X_t + \theta Y_t, \tag{4.31}$$

where X_t is a vector of observable "taste shifter" controls and Y_t is a measure of non-labor income. Non-labor income is typically measured as the sum of interest income $r_t A_{t-1}$ and exogenous income B_t. This static specification is only appropriate if the static model of Section 4.1 is correct. This could be the case if consumers behave completely myopically or if capital markets are completely constrained so that it is impossible to transfer capital across periods. If the static model is correct, the wage coefficient in specification (4.30) measures

$$\alpha = \alpha_y = \text{uncompensated substitution elasticity given income } Y. \tag{4.32}$$

The parameter α_y corresponds to the Marshallian wage elasticity in the static model. Its estimation requires instrumental-variable techniques to account for the endogeneity of the wage, arising from unobservable characteristics affecting both W_t and H_t or from measurement error. Nevertheless, if consumers adjust their behavior to account for factors in future periods, the coefficient on log wage lacks economic meaning, no matter what econometric methods are applied. That is, if the labor supply decision has any lifecycle elements, static regressions confuse shifts of wage profiles with movements along wage profiles and, thus, yield parameters that lack economic interpretation.

4.4.2. Two-stage budgeting specifications

To estimate a labor supply equation within a two-stage budgeting framework, set

$$\beta Q_t = \rho X_t + \theta Y_t^C, \tag{4.33}$$

where Y_t^C is the consumption-based income measure defined in (4.20′). (Alternatively, one can condition on the full income measure M_t defined by (4.20)). In applying these controls, one should note that Y_t^C (or M_t) is defined by leisure and consumption choices and, thus, is endogenous. Appropriate instrumental-variable techniques must, therefore, be applied to obtain consistent estimators. The wage elasticity coefficient, α, can then be interpreted as

$$\alpha = \alpha_c = \text{uncompensated substitution elasticity given total consumption } C \tag{4.34}$$

This wage effect determines the impact of wages on hours worked, holding the first-stage income allocation constant. Hence, it captures the impact of anticipated wage movements through time, but does not capture the impact of shifts of the entire wage profile, as these shifts would also impact hours through their effect on the allocation of Y_t^C. In general, one needs a model of Y_t^C that includes the impact of all current and future wages to assess the impact of wage profile shifts. We take up this issue further in the discussion of relevant elasticities for policy evaluation in Section 4.5.

4.4.3. Frisch specifications

To create a Frisch labor supply function in the form of (4.30), suppose the contemporaneous utility function for period t takes the form

$$U_t = G(C_t, X_t) - \Psi_t(H_t)^\sigma, \tag{4.35}$$

where G is a monotonically increasing function of C_t, $\sigma > 1$ is a time-invariant parameter common across consumers and Ψ_t is a function of consumer characteristics. We take Ψ_t to be $\exp(-X_t\rho^* - \nu_t^*)$ where ν^*_t reflects the contribution of unmeasured characteristics and ρ^* is a vector of preference parameters.

Assuming an interior optimum, the implied Frisch hours-of-work function takes the form of (4.30) with

$$\beta Q_t = F_t + \rho X_t, \tag{4.36}$$

where $F_t = \alpha(\ln\lambda_t - \ln\sigma)$, $\alpha = 1/(\sigma - 1)$, $\rho = \alpha\rho^*$, and $e_t = \alpha\nu_t^*$. Modifying Eq. (4.29) by assuming that the b^*_t terms are constant across consumers and time, and substituting this into (4.36) yields

$$\beta Q_t = F_0 + bt + \rho X_t, \tag{4.37}$$

where $b = \alpha b^*$, and e_t now includes sums of forecast error terms. So, the necessary controls are the exogenous variables X_t, age and an individual effect F_0. Taking first differences of this form of Eq. (4.30) yields

$$\Delta\ln H_t = b + \rho\Delta X_t + \alpha\Delta\ln W_t + \Delta e_t. \tag{4.38}$$

Given the availability of instruments for the change in wage, one can fit this equation on panel data to yield an estimate of α. In these specifications α corresponds to the Frisch wage elasticity discussed above, which the literature commonly designates

$$\alpha = \alpha_t = \text{intertemporal substitution elasticity}. \tag{4.39}$$

This elasticity holds marginal utility of wealth constant, and it describes how changes in wages induced by movements along an individual's wage profile influence hours of work. Individuals fully anticipate these wage movements and this is why F_0 remains fixed. For this reason, they are often referred to as evolutionary wage changes.[14]

If we wish to measure the impact of wage variation across consumers, or unanticipated shifts of an individual's wage profile, we must complete the model and provide an empirical specification of the evolution of wages and other incomes as well as accounting for the impact of these shifts on F_0. Hence, we need an empirical specification for λ and, thus, for F. This is provided by the lifecycle specifications that we now consider.

4.4.4. Lifecycle specifications

For this empirical specification, we assume that one can approximate $\ln\lambda_0$ by the equation

$$\ln\lambda_0 = D_0\varphi_0^* + \sum_{j=0}^{\tau}\gamma_{0j}^* E_0\{\ln W_j\} + \theta_0^* A_0 + a_0^*, \tag{4.40}$$

[14] Although the particular form for utility, (4.35), conveniently implies a log-linear Frisch labor supply equation, it also places strong restrictions on the form of within period and intertemporal preferences. In this specification, labor supply and consumption are explicitly additive in utility both within period and across periods.

where D_0 is a vector of demographic characteristics either observed at 0 or anticipated in future periods, and a^*_0 is an error term. This implies a form for F_0:

$$F_0 = D_0\varphi_0 + \sum_{j=0}^{\tau} \gamma_{0j} E_0\{\ln W_j\} + \theta_0 A_0 + a_0, \tag{4.41}$$

where the parameters and error term equal their superscript "*" counterparts multiplied by α, and with the intercept defined to include the term $-\alpha \ln\sigma$. This empirical specification imposes strong simplifying restrictions – it assumes that the consumer knows he will work τ periods and it incorporates any effect of interest rates or time preference into the intercept and other parameters.

Relations (4.41) and (4.37) yield a formulation for (4.30) with

$$\beta Q_t = D_0\varphi_0 + \sum_{j=0, j\neq t}^{\tau} \gamma_{0j} E_0\{\ln W_j\} + \theta_0 A_0 + bt + X_t\rho, \tag{4.42}$$

$\alpha = \alpha_t + \gamma_{0t}$, where the disturbance in (4.30) is $e_t = a_0 + \nu_t - \gamma_{0t}(\ln W_t - E_0\{\ln W_t\})$. So, Q_t now includes all start-of-life controls used to form λ_0 and all controls needed for the period-t utility function: age, initial wealth and the expected wage profile as of age 0. Estimation of this equation yields an estimate of $\alpha_t + \gamma_{0t}$, the wage elasticity of hours corresponding to a shift in the period t wage rate, as well as estimates of the γ_{0t}s determining the impact of a shift in the entire wage profile. As we argue below, this formulation also provides us with precisely the parameters we need for the analysis of tax reform.

Implementing (4.42) requires the econometrician to have consistent predictions of the consumer's expected future wages. Assume that the lifetime wage path anticipated in period 0 is

$$E_0\{\ln W_t\} = \pi_o + \pi_1 t + \pi_2 t^2 + u_t, \tag{4.43}$$

where the π's are deterministic functions of time invariant characteristics of the consumer and u_t is an error term assumed to be uncorrelated with all demographic variables in M_0 as well as with those used to predict wages or wealth (below).

A researcher also requires a specification for initial wealth since most datasets do not include this variable. If we assume that property income, Y_t, follows a path similar to wages (with the similar properties for errors and parameters)

$$E_0\{Y_t\} = \zeta_o + \zeta_1 t + \zeta_2 t^2 + \eta_t, \tag{4.44}$$

then using the fact that $Y_0 = (A_0/1 + r_0)r_0$, we see that initial wealth can be predicted by $\zeta_0(1 + r_0)/r_0$.

Combining these forms for wages and wealth with (4.41), we arrive at an expression for the individual effect:

$$F_0 = D_0\varphi_0 + \pi_0\bar{\gamma}_0 + \pi_1\bar{\gamma}_1 + \pi_2\bar{\gamma}_2 + \zeta_0\ddot{\theta} + \mu, \tag{4.45}$$

where

$$\bar{\gamma}_k = \sum_{j=0}^{\tau} j^k \gamma_{0j}, \qquad \text{for } k = 0, 1, 2, \qquad \bar{\theta} = \theta_0 r_0/(1 + r_0)$$

and μ is a disturbance depending on the errors a_0, u_t's and η_t's. This equation relates a consumer's individual effect to the parameters of his wage and income profile. Relations (4.37) and (4.45) yield a formulation for (4.30) with

$$\beta Q_t = D_0\varphi_0 + \pi_0\bar{\gamma}_0 + \pi_1\bar{\gamma}_1 + \pi_2\bar{\gamma}_2 + \zeta_0\bar{\theta} + bt + X_t\rho, \tag{4.46}$$

$$\alpha = \alpha_t + \gamma_{0t},$$

where the disturbance e_t in (4.30) now incorporates the error component μ. Hence, in this formulation, D_0 and age remain as controls, but initial wealth and the expected wage profile are replaced by the parameters describing wage and property income profiles through time. Simultaneous estimation of (4.43), (4.44) and (4.46) yields estimates of all parameters needed to compute the response of hours of work to both evolutionary and parametric wage changes. In this formulation, only wages and property income are endogenous.

4.4.5. Interpreting cross-sectional specifications in a lifecycle framework

Many labor supply studies attempt to estimate "wage elasticities" using cross-sectional variation in wages. As we have seen above, the term wage elasticity is ambiguous – it is crucial that the researcher distinguish between evolutionary and parametric wage shifts. Since most do not, the reader is left trying to compare elasticity estimates that may not be comparable. Add to this the difficulty of identifying any lifecycle effects in a cross-sectional setting and, even if there were no data measurement differences, it would not be surprising to see many different elasticity estimates.

Utilizing the above framework, we can evaluate what cross-sectional specifications actually allow meaningful lifecycle parameter estimates to be recovered and which specific parameters are being estimated given the included control variables. To develop a simple expression for (4.46) which can be compared to those of existing cross-sectional studies, assume that $D_0\phi_0$, π_0, π_1, π_2, and ζ_0 are linear functions of the variables contained in a vector, K. Then we have

$$\ln H_t = Kq + bt + X_t\rho + (\alpha_t + \gamma_{0t})\ln W_t + e_t, \tag{4.47}$$

where q is a vector of coefficients. Alternatively, we could assume that D_0 contains only an intercept and that the coefficients on age and age-squared for the lifetime wage and income paths (i.e., π_0, π_1, π_2, and ζ_0) are constant across consumers. Then one can write (4.46) as

$$\ln H_t = d_1 + d_4 t + d_5 t^2 + \bar{\theta} Y_t + X_t\rho + (\alpha_t + \bar{\gamma}_0)\ln W_t + e_t, \tag{4.48}$$

where

$$d_1 = g_0 + \pi_1\bar{\gamma}_1 + \pi_2\bar{\gamma}_2,$$

$$d_4 = b - \pi_1 \bar{\gamma}_0 - \gamma_2 \zeta_1,$$

$$d_5 = -\pi_2 \bar{\gamma}_0 - \theta \alpha_2 \zeta.$$

So, there are two equations which one can estimate using instrumental variable techniques on cross-sectional data to yield meaningful lifecycle parameter estimates. If a researcher regresses log hours of work on age; all age-invariant characteristics determining lifetime wages, preferences, and initial permanent income; and log wage, then the coefficient on the current wage rate is α, the Frisch elasticity. Intuitively, this approach controls for differences in the value of F_0 across consumers and leaves higher-order age variables as instruments to identify wage variation. Hence, only evolutionary wage variation along the age-wage path is included.

If, alternatively, a researcher regresses log hours worked on property income, age, age squared, and log wage, the coefficient on wage is the response of labor supply to a parametric wage shift – including both the intertemporal substitution effect, α, and the reallocation of wealth across periods captured by a change in F. Intuitively, this approach controls for age effects and leaves individual characteristics as instruments for wage. Changes in these characteristics capture full profile shifts, rather than movements along the age-wage path. The static equations presented in (4.30) fit neither of these patterns, however, as they include property income together with personal characteristics rather than age and age squared. Hence, as noted above, given the existence of lifecycle effects they confuse the effect of movements along the wage profile with shifts in the profile and, thus, yield parameters without an economic interpretation.

4.5. Which elasticities for policy evaluation?

This section has highlighted four "core" wage elasticities which correspond to four key specifications for control variables that can be found in the empirical literature on labor supply. Two are within-period elasticities: α_Y relating to the purely static formulation (4.31) and α_C relating to the two-stage budgeting specification (4.33). Two are lifecycle elasticities: α_l the intertemporal elasticity of substitution relating to the Frisch specification (4.38) and measuring responses to evolutionary movements along the lifecycle wage profile, and $\alpha_l + \gamma_0$ relating to the full lifecycle specification (4.42) and measuring responses to parametric shifts in the lifecycle profile itself. As most tax and benefit reforms are probably best described as once-and-for-all unanticipated shifts in net-of-tax real wages today and in the future, the most appropriate elasticity for describing responses to this kind of shift is $\alpha_l + \gamma_0$. Here we examine the relations among each of the elasticities and consider their relevance for policy evaluation.

4.5.1. Relationships among the lifecycle elasticities

The Frisch specification treats the individual marginal-utility-of-wealth as a fixed effect and allows the researcher to estimate only the intertemporal substitution elasticity, α_l. Given that appropriate methods are employed to account for the fixed effect (generally first

differencing in panel data), the relevant independent variables, apart from the wage, are simply within-period characteristics and age.[15] The Frisch elasticity, by ignoring this (unexpected) shift in wealth from a once-and-for-all change in real wages, is larger than the policy-relevant elasticity $\alpha_I + \gamma_0$ and overestimates the impact of a reform.

Direct estimation of the simple parameterization of the full lifecycle model, required to recover $\alpha_I + \gamma_0$, relies on specifications for both within-period utility and the individual marginal-utility-of-wealth effect. As a result, controls are needed for all of the following: "start of life" characteristics which impact the initial setting of F_0, current-period characteristics which affect the within-period utility function, age, expected wages as of time 0, and initial wealth. Expected wages are unobservable and initial wealth is generally not included in data sets, so these should be replaced with the parameters governing the time path of wages and property income, which must be jointly estimated with the labor supply equation. Estimation of this full framework allows computation of both the intertemporal substitution elasticity and the elasticity of labor supply in reaction to a full, parametric wage profile shift. However, it is also the most demanding in terms of data.

It is worth noting that the elasticity derived from the static specification, α_γ, can be placed in an intertemporal setting but is economically meaningful only under a strong assumption of either complete myopia or perfectly constrained capital markets. Otherwise, this elasticity confuses movements along wage profiles with shifts of these profiles and, thus, yields response parameters which are a mixture of these. Such hybrid estimates lack an economic interpretation and are not generally useful in policy evaluation.

However, we have also described several formulations which appear essentially static, but which vary greatly based on included controls. Under simplifying assumptions, formulation (4.47) allows the researcher to compute the intertemporal substitution elasticity using cross-sectional data alone. Age and age-invariant consumer characteristics are the required controls. In contrast, formulation (4.48) allows one to estimate the response to a parametric wage shift. Required controls here are property income in period t, age, and age squared.

4.5.2. Relationships among within-period and lifecycle elasticities

In general, a tax policy reform will lead to a change in the optimal level of consumption and full income. The within-period elasticity, α_C, based on the two-stage budgeting framework, does not account appropriately for intertemporal adjustments in consumption. So how should we interpret elasticity α_C from the two-stage budgeting formulation? Under the strong assumption of either complete myopia or perfectly constrained capital markets, this elasticity is identical to α_γ. But in the lifecycle model with capital markets, the precise relationship between the policy-relevant elasticity, $\alpha_I + \gamma_0$, and α_C is ambiguous. However, since α_C is bounded above by the Slutsky compensated elasticity and α_I is

[15] In the model with uncertainty the fixed effect is replaced by a random walk (see Section 4.3.2), but the first difference solution to estimation is retained with appropriate adjustment for the endogeneity of differenced wages.

bounded below by the Slutsky elasticity, α_C is no greater than the Frisch elasticity. It may well be much smaller and, unlike α_I, can be negative.

Indeed, in certain cases, α_C precisely reflects the labor supply adjustment induced by the shift in wealth, capturing exactly the impact of the parametric shift in the wage profile that corresponds to a policy reform involving an unexpected and permanent change in real wages. To see this, consider the case where within-period preferences are of Stone–Geary form

$$U_t = \theta \ln(\gamma_H - H_t) + (1 - \theta)\ln(C_t - \gamma_C), \tag{4.49}$$

where θ, γ_H and γ_C are preference parameters. Suppose also that intertemporal preferences are explicitly additive over U_t.[16] The labor supply specification from the two-stage budgeting approach has the form

$$H_t = \gamma_H - (\theta/W_t)\{Y_t^C - \gamma_C + \gamma_H W_t\} \tag{4.50}$$

and the within-period elasticity is

$$\alpha_C = \frac{W}{H}\cdot\frac{\partial H}{\partial W}\bigg|_{Y^C} = \frac{\gamma_H}{H}(1 - \theta) - 1. \tag{4.51}$$

To compare this elasticity with $\alpha_I + \gamma_{0t}$, we can compute the following expression for λ_t^{-1}:

$$\lambda_t^{-1} = A_{t-1} + \sum_j (\kappa(1 + r))^{-j}(\gamma_H W_j - \gamma_C). \tag{4.52}$$

Now consider a permanent change in the wage, W. Assume (i) $\kappa(1 + r) = 1$ and (ii) future real wages remain at this new level. The corresponding elasticity is

$$\frac{W}{H}\frac{\partial H}{\partial W} = \frac{\gamma_H}{H}(1 - \theta) - 1, \tag{4.53}$$

which, in this case, is identical to the within-period uncompensated elasticity from the two-stage budgeting formulation (4.51). In this case, it turns out that the consumption-based measure of other income, Y_t^C, is constant for a permanent uniform shift in real wages and, consequently, α_C matches the policy-relevant elasticity. Consumption levels adjust but are exactly offset by the change in $W_t H_t$ in the definition of $Y_t^C = C_t - W_t H_t$. This example shows that, in certain cases, the adjustment for the wealth effect needed to account for the unexpected and permanent change in future wages arising from a policy change is completely captured in the two-stage budgeting formulation. It also highlights the degree to which the intertemporal substitution elasticity overestimates the policy relevant effect.

For completeness, consider now the Frisch elasticity for this Stone–Geary specification. The Frisch labor supply has the form

[16] See Ashenfelter and Ham (1979) and Bover (1989) for further discussion of this specification.

$$H_t = \gamma_H - (\theta/W_t)\lambda_t^{-1}, \tag{4.54}$$

with elasticity given by

$$\alpha_I = \frac{W}{H}\frac{\partial H}{\partial W}\bigg|_{\lambda} = \frac{\gamma_H}{H} - 1. \tag{4.55}$$

This intertemporal substitution elasticity must be non-negative since $\gamma_H \geq H_t$ and, since θ lies between zero and one, this elasticity is larger than α_C from the two-stage budgeting formulation.

In general, the equivalence between α_C and $\alpha_I + \gamma_0$ found in this Stone–Geary example without uncertainty does not hold. The Stone–Geary preference specification and explicit additivity over time places strong restrictions on preferences. In Appendix A we describe the properties of this and other popular preference models for within-period labor supply.

One general way to exploit the simplicity of the second stage of the two-stage budgeting formulation under uncertainty is to use the linkages between within-period and intertemporal preference restrictions.[17] This combines the within period stage, which conditions on consumption, with an Euler equation for the marginal utility of wealth under uncertainty. All preference parameters needed to describe both stages of the intertemporal allocation model under uncertainty are identified by combining the second stage of the two-stage budgeting framework with the Euler equation for consumption. Within-period allocations between consumption and leisure are completely described by the labor supply equations that condition on the consumption-based measure of full income or the marginal rate of substitution condition between consumption and hours. The Euler condition on the marginal utility of wealth then recovers the remaining parameters describing intertemporal allocations.

This approach of combining the two-stage budgeting formulation with the Euler equation for the marginal utility of consumption has many potential advantages over the Frisch and full lifecycle approaches. Frisch labor supply models specify hours of work directly in terms of wages and the marginal utility of wealth. The strong restrictions on preferences in the standard log linear specification can be seen directly from the implied form of utility in (4.35). Utility is explicitly additive over time, goods and leisure. In general, for the Frisch labor supply model to be log linear in the wage and log marginal utility, the intertemporal utility must be explicitly additive over time, consumption and hours. However, the two-stage budgeting approach requires accurate measurement of consumption as well as labor supply and real wages. Moreover, there are many potential pitfalls. Additive heterogeneity

[17] Consider writing the period-specific utility function $U(C_t, L_t, X_t)$ in the intertemporal program (4.27) as $U(C_t, L_t, X_t) = G(u(C_t, L_t, X_t), X_t)$, where G is some positive monotonic transformation of a quasi-concave, differentiable, within-period utility, u. This expression is convenient since the marginal within-period allocation conditions (4.4) become $G_u u_t(C_t, L_t, X_t) = \lambda_t$ and $G_u u_L(C_t, L_t, X_t) \geq \lambda_t W_t$. The marginal rate of substitution $u_L/u_C \equiv MRS_L(C_t, L_t, X_t) \geq W_t$ does not depend on G. Consequently, within-period allocations place no restrictions on G and, therefore, provide no information on the identification of G. In contrast, the Euler condition (4.28) involves the derivatives of G and u. Given u, the form of G places restrictions on intertemporal preferences.

at the within-period level does not fit easily into a non-linear Euler equation. Similar issues arise with measurement error, endogeneity and non-participation.[18] As with static labor supply models, simple specifications may be preferred in empirical applications where heterogeneity and measurement error are considered to be overriding issues.

4.5.3. Summary and some qualifications

This section has demonstrated the importance of understanding which elasticity is being recovered in the empirical analysis of labor supply and has shown that this depends crucially on the conditioning variables included in estimation. We have identified four "core" elasticities that are commonly estimated and which differ substantially in their interpretation. For this purpose we have abstracted, in this section, from important issues such as non-linear taxation, discreteness in choices and flexibility in the specification of preferences, so as to highlight the differences in interpretation of coefficients across alternative specifications. We have argued that, in general, a full lifecycle parameterization of the model is needed to evaluate policy reforms. However, we have shown how key policy-relevant elasticities can be recovered from the analysis of available data sources.

The analysis presented here and elsewhere in this chapter is conducted in a partial equilibrium framework and, therefore, considers only one side of the market. To analyze the impact of a policy reform, a general equilibrium analysis will sometimes be required, though discussion of this is outside the scope of this chapter. The model specifications examined in this section have been stylized and often relate to simple linear formulations, which place strong restrictions on preferences.[19] Furthermore, in focusing on one side of the market, these specifications may not directly capture short-term constraints on the adjustment of labor supply. Nevertheless, they do include error terms to reflect this and should be viewed as representing "average" behavior. Ham (1986a,b) provides evidence of the importance of short-run constraints. Extreme liquidity constraints may also limit the usefulness of the intertemporal model. Finally, it may be that these simple intertemporal models are inappropriate for certain types of workers. For example, in a (unionized) bargaining model, hours-wages contracts might implicitly allow for smoothing consumption via clauses that provide for a steady stream of income in exchange for additional effort from the workers in good times. See Card (1994) for a critical review of the intertemporal labor supply model. In Section 8 we consider many extensions of the basic intertemporal model, though we focus only on those extensions that allow for human capital, habits and discrete participation choices.

5. Policy reforms and the natural experiment approach

"Natural experiments" have gained considerable popularity recently, and the simplicity of

[18] Section 8 considers the introduction of participation in this formulation.

[19] In Appendix A, we summarize the preference restrictions underlying popular parameterizations of labor supply.

this estimation method will undoubtedly make its popularity enduring among empirical economists for some time to come. This method often goes by the name of the difference-in-difference estimator. This section interprets the essence of this approach, and it relates those applications that estimate how tax and welfare policies influence labor supply to the empirical models surveyed elsewhere in this chapter. Although the discussion focuses on labor supply analyses, the evaluation presented here applies to any implementation of the natural-experiment approach.

The natural-experiment approach is not new, nor is it a method that is "non-structural". The statistical apparatus underlying this approach has been extensively applied in the labor-economics literature since the inception of empirical work in the field. The basic idea is to compare (at least) two groups, one of which experienced a specific policy change, and another with similar characteristics whose behavior was unaffected by this policy change. The second group is assumed to mimic a control environment in experimental terminology. Such comparisons provide the foundation for most empirical work in labor economics. The problem comes in creating a control environment, which is done either by including exogenous variables in an analysis designed to adjust for relevant differences among sample observations, or by selecting observations in a manner that permits a matched-pair type of analysis.

Contrary to many researchers' perceptions, the natural experiment approach relies on restrictive structural assumptions analogous to those of most other methods. In fact, this approach is entirely equivalent to the fixed-effects model popularized in the 1970s. By writing the model in this way, we are able to compare it with the alternative structural models outlined in the previous section and to state the conditions under which a structural interpretation can be placed on estimates from studies that use this approach.

5.1. The natural-experiment approach and the difference-in-differences estimator

Suppose one is interested in estimating the influence of a policy instrument on an outcome for a group, say outcome y_{it} measuring hours of work or participation. The group consists of individuals $i = 1,\ldots,N$, with these individuals observed over a sample horizon $t = 1,\ldots,T$. (Individuals here may refer to data on groups such as the average in a state or in a specific demographic category.) Suppose further that the policy instrument changes in particular period t for only a segment of the group. Let δ_{it} be a zero-one indicator that equals unity if the policy change was operative for individual i in period t. Members of the group who experience the policy change react according to a parameter γ. A framework for estimating γ expressed in terms of a conventional fixed-effect model takes the form

$$y_{it} = \gamma\delta_{it} + \eta_i + m_t + \varepsilon_{it}, \tag{5.1}$$

where η_i is a time-invariant effect unique to individual i, m_t is a time effect common to all individuals in period t, and ε_{it} is an individual time-varying error distributed independently across individuals and independently of all η_i and m_t.

Estimation of coefficients in "error-components" models, of which Eq. (5.1) is a special

case, occupies an extensive econometrics literature. Balestra and Nerlove (1966) and Nerlove (1971) discuss a variety of estimation procedures under various assumptions regarding the distributions of η_i and m_t. When η_i and m_t are random components, meaning their distributions are independent of observed right-hand side variables, then conventional generalized least squares produces an estimator that is consistent and asymptotically efficient.[20] When the distributions of η_i and m_t depend on right-hand side variables, the literature implements a differencing procedure to calculate consistent estimators, where the form of differencing depends on the particular nature of the simultaneity problems induced by η_i and m_t. Analysts commonly refer to these as "within" estimators because they rely only on variation within groups in calculations. The fixed-effect estimator, which treats η_i and m_t as parameters, is a special case of such an estimator.

5.1.1. Difference-in-differences estimators

Suppose both η_i and m_t are believed to be dependent on δ_{it} in some unknown manner, and one wants to compute a consistent estimate of γ in (5.1). A popular version of a within estimator involves first differencing (5.1) over time to obtain

$$\Delta_t y_{it} = \gamma \Delta_t \delta_{it} + \mu_t + \Delta_t \varepsilon_{it}, \tag{5.2}$$

where $\Delta_t y_{it} \equiv y_{it} - y_{i(t-1)}$ and $\mu_t \equiv \Delta_t m_t$. The operator Δ_t differences an individual's observation across periods, and μ_t is merely defined to be a parameter representing the difference in common time effects.

Suppose, for simplicity, that the sample consists of only two periods: period $t-1$ which is before the implementation of the policy instrument and period t which is after. Let group e represent the "experimentals", the individuals who experienced the change in the policy instrument – and let group c denote the "controls" – the individuals who encountered no policy change. Then least squares applied to (5.2) yields the estimators

$$\hat{\gamma} = \Delta_t \bar{y}^e - \Delta_t \bar{y}^c, \qquad \hat{\mu} = \Delta_t \bar{y}^c, \tag{5.3}$$

where

$$\Delta_t \bar{y}^k = \bar{y}_t^k - \bar{y}_{t-1}^k, \qquad k = e, c,$$

$$\bar{y}_j^k = \frac{\sum_{i \in k} y_{ij}}{N_k}, \qquad k = e, c,$$

where $\bar{y}_j^k$ is the average outcome for group k.[21]

The estimator $\hat{\gamma}$ in (5.3) is identical to what is now known in the literature as the

[20] This estimator accounts for the autocorrelation implied by the disturbance $\eta_i + m_t + \varepsilon_{it}$ for an individual, and for correlation across individuals implied by the disturbances m_t.

[21] The notation $\sum_{i \in k}$ designates that summation is over all individuals included in group k, and N_k is the total number of individuals in group k.

difference-in-difference estimator. The fixed-effect and difference-in-difference estimators do not merely share the same asymptotic distribution; they are computationally identical.

The literature considers many generalizations of fixed-effects models, which in turn imply generalizations of the natural-experiment approach. A common extension incorporates covariates in (5.1) to obtain

$$y_{it} = \gamma\delta_{it} + Z_{it}\theta + \eta_i + m_t + \varepsilon_{it}, \tag{5.4}$$

where Z_{it} includes observed exogenous and/or endogenous variables.[22] A further generalization of this model allows for treatment effects to vary randomly across individuals. Under the stringent structural assumptions on time effects and composition highlighted below, the difference-in-differences estimator can be shown to recover the average treatment effect for the treated (i.e., the parameter $E(\gamma \mid \delta_{it} = 1)$). Unfortunately, this parameter is subject to conventional sample selection biases and in general cannot be used to simulate policy responses.

5.1.2. Structural assumptions maintained by the difference-in-difference estimator

Applications of the natural experiment approach typically suggest that it is a "non-structural" estimation procedure, but its equivalence to error-components models clearly indicates that all of the restrictions required for consistent estimation of these models must also hold for the difference-in-difference estimator to measure a behaviorally meaningful parameter. The literature has never interpreted the fixed-effect model as non-structural. The requirement of two sets of structural restrictions are likely to challenge the credibility of many natural-experiment applications concerned with estimating behavioral responses in labor supply.

Assumption 1. Time effects in (5.1) (or (5.4)) must be common across experimentals and controls.

More flexible specifications of (5.1) include the following:

$$y_{it} = \gamma\delta_{it} + \eta_i + m_{ct} + m_{et} + \varepsilon_{it} \tag{5.5}$$

and

$$y_{it} = \gamma\delta_{it} + \lambda_t\eta_i + m_t + \varepsilon_{it}. \tag{5.6}$$

Many factors can lead to these generalizations, including failure to include relevant time-varying variables in Z_{it} that differ across experimentals and controls. Specification (5.5) recognizes that experimentals and controls might experience dissimilar trends and/or cyclical effects. Such an event is likely, for example, when the demographic composition of experimentals and controls differs; empirical analysis usually shows that the trends and

[22] Hausman and Taylor (1981) and Amemiya and MaCurdy (1986), for example, develop asymptotically efficient estimators for model (5.4).

cycles differ for married and single people, for men and women, and for high- and low-skilled workers. Specification (5.6) allows individual effects to influence outcomes differentially over time. This phenomenon often happens in analyses of work or wage outcomes over the life cycle. An analysis of the differential time trends, before and after the policy intervention, for each group provides useful information in assessing the reliability of this assumption.

Assumption 2. The composition of both experimentals and controls must remain stable before and after the policy change.

The averages in (5.3) presume that the same individuals make up each group in both period t and period $t-1$. If this is not the case, then differencing does not eliminate averages of the individual effects η_i. Instead, the terms

$$\Delta_t \eta^e = \bar{\eta}^{e_t} - \bar{\eta}^{e_{t-1}}, \qquad \Delta_t \eta^c = \bar{\eta}^{c_t} - \bar{\eta}^{c_{t-1}},$$

with

$$\bar{\eta}^{k_j} = \frac{\sum_{i \in k_j} \eta_{ij}}{N_{k_j}}, \qquad k_j = e_t, e_{t-1}, c_t, c_{t-1}$$

contaminate the estimate of γ given by (5.3). Even when the groups e_t and e_{t-1} consist of different individuals, it can still happen that $\Delta_t \eta^e$ vanishes asymptotically keeping $\hat{\gamma}$ consistent. These circumstances typically involve random selection mechanisms. However, selection into groups made up of workers, as is the case in analyses of labor supply, is invariably not random since it depends intricately on the nature of the policy change. For example, a tax change can be expected to alter who works and who does not in a systematic manner. As a consequence, sample selection terms prevent $\Delta_t \eta^e$ from vanishing. Exactly the same problem arises for a shifting composition of the control groups c_t and c_{t-1}, which keeps $\Delta_t \eta^c$ from disappearing.

5.1.3. Grouping estimators

Applications occasionally have grouped data available for their analyses, or they may have a discrete grouping variable (instrument) G_{it} that allocates individuals into $g = 1,\ldots,J$ groups of size N_{gt} in each period $t = 1,\ldots,T$. A modest modification of fixed-effect model (5.1) (or (5.4)) provides a framework for estimating relevant coefficients in many of these cases. Suppose also that the discrete grouping variable satisfies the assumption

$$y_{it} = \gamma \delta_{it} + \theta_g + \eta_i + m_t + \varepsilon_{it}, \tag{5.7}$$

where θ_g is a time-invariant effect unique to group g and η_i is now an error reflecting the deviation of a particular observation's individual effect around its respective group mean. Defining the group averages

$$\bar{y}_{gt} = \frac{\sum_{i \in g} y_{it}}{N_g}, \qquad \bar{\delta}_{gt} = \frac{\sum_{i \in g} \delta_{it}}{N_g}, \qquad \bar{\varepsilon}_{gt} = \frac{\sum_{i \in g} \varepsilon_{it}}{N_g},$$

and averaging Eq. (5.7) over groups yields

$$\bar{y}_{gt} = \gamma \bar{\delta}_{gt} + \theta_g + m_t = \varepsilon_{gt}. \tag{5.8}$$

This is just another version of a fixed-effect model, as long as one maintains the structural assumptions for the error components θ_g, m_t and $\bar{\varepsilon}_{gt}$ for grouped data analogous to those outlined in Section 5.1.2 for the components η_i, m_t, and ε_{it} using individual data.

Estimation of model (5.8) – or its variant with the grouped covariates $\bar{X}_{gt}$ also included – involves no complications beyond those already discussed.[23] Under the structural assumptions presumed for the conventional fixed-effect model, differencing eliminates the source of endogeneity for $\bar{\delta}_{gt}$. The quantity $\bar{\delta}_{gt}$ represents the proportion in group g receiving the treatment. The asymptotically efficient estimators developed for model (5.4) apply here as well, with instrumental variables now specified for groups. When there are two groups and when the grouping instrument coincides exactly with the policy reform dummy variable δ_{it}, then this estimator is identical to the difference-in-differences estimator. In any particular application, the objective is to find a suitable grouping instrument such that the resulting grouped error components satisfy the structural conditions of the fixed-effect specification.

5.1.4. Repeated cross-section or panel data?

Since the difference-in-differences estimator and the instrumental variable estimator defined by Eq. (5.3) are expressed in terms of sample means, they can be computed equally well using either repeated cross-section or panel data. Panel data only become useful when the instrumental variable method uses an historic individual variable as an instrument. For example, if past employment status or past tax status is the instrument, then this estimator would typically not be available using cross-section data.

In both the panel data and the repeated cross-section case, the structural conditions are still needed to pursue the difference-in-difference estimator. Provided there is no systematic attrition across groups, panel data allow the groups to be determined in a time-invariant way and, therefore, the difference-in-difference approach completely eliminates the individual fixed effects η_i. Thus, no restrictions need be placed on the distribution of the individual effects. Repeated cross section data, on the other hand, must satisfy the assumption that the unobservable individual effects are drawn from the same population distribution across periods before and after the reform. Otherwise, the difference-in-differences estimator and the instrumental-variable estimator suffer from composition bias. Panel data applications still require the strong restrictions on the distribution of the individual "transitory" time-varying effects and must retain the common-trend assumption.

[23] See also Angrist (1991) and Moffitt (1993).

5.2. Does the difference-in-differences estimator measure behavioral responses?

Most advocates of the natural-experiment approach would answer this question as NO, and they would be right if behavioral responses refers to substitution and income effects familiar in labor supply analyses. Indeed, researchers applying a difference-in-difference procedure often emphasize that they have no intention of estimating such effects.

What, then, is the interpretation of γ in Eq. (5.1) (or Eq. (5.4))? Clearly, under ideal circumstances, γ measures the total response of a policy change, or, more precisely, how a shift in a policy regime influences the average outcome for a worker in the experimental group. But one can seldom translate this response into interpretable behavioral effects because most shifts in policy regimes involve simultaneous changes in marginal wages and net income, and rarely are these changes the same for all individuals making up a group.

To illustrate the issues, reconsider the prototype empirical specification given by Eq. (4.30), which we repeat here for convenience:

$$\ln H_{it} = \alpha \ln W_{it} + \beta Q_{it} + e_{it}. \tag{5.9}$$

Suppose that a policy shift results in changes in $\ln W$ and Q equal to $\Delta_t \ln W$ and $\Delta_t Q$, respectively. A translation of this model into the simple fixed-effect framework,

$$y_{it} = \gamma \delta_{it} + \eta_i + m_t + \varepsilon_{it},$$

is possible by specifying

$$y_{it} = \ln H_{it}, \tag{5.10a}$$

$$\gamma = \overline{\alpha \Delta_t \ln W_{it} + \beta \Delta_t Q_{it}}, \tag{5.10b}$$

$$\eta_i + m_t = \alpha \ln W_{it} + \beta Q_{it}, \tag{5.10c}$$

$$\varepsilon_{it} = e_{it} + \delta_{it}[\alpha \Delta_t \ln W_{it} + \beta \Delta_t Q_{it} - \gamma]. \tag{5.10d}$$

The coefficient γ is the average of $\alpha \Delta_t \ln W_{it} + \beta \Delta_t Q_{it}$ among the experimentals, and the error ε_{it} includes the difference between $\alpha \Delta_t \ln W_{it} + \beta \Delta_t Q_{it}$ and its mean as one of its components. This is one interpretation of the heterogeneous treatment-effects model discussed in Section 5.1.1. Formulation (5.9) assumes that only experimentals experience the change in policy, with $\delta_{it} = 1$ signaling the periods and individuals affected by the change. Specifications (5.9) restrict the permissible variation in W and Q across individuals and time; a form of variation satisfying this property occurs when both $\ln W_{it}$ and Q_{it} can be represented as the sum of an individual and time effect. (Of course, consideration of fixed-effect formulation (5.4) permits some relaxation of these variability restrictions.) The natural-experiment framework requires ε_{it} to be independent of δ_{it}, meaning neither the structural error, e_{it}, nor changes in W and Q provide any information indicating whether an individual is in the experimental group or not.

In this idealized model, the difference-in-difference estimator for γ measures a weighted substitution-income effect given by Eq. (5.10b). We know from our discussion in Section 4 that the interpretation of this combined effect depends on the other control variables included in Q. If one imagines a situation in which Q properly includes a measure of static income or within-period expenditure (such as Eqs. (4.31) or (4.33)), then the substitution effect α corresponds to an uncompensated substitution elasticity. If, on the other hand, Q incorporates age-invariant characteristics controlling for lifetime wages, preferences and initial permanent income (see Section 4.6), then α conforms to the intertemporal substitution elasticity. For still another interpretation, if Q now includes controls for age, initial wealth and the expected wage profile (such as Eq. (4.46)), then α measures the wage elasticity of hours corresponding to a shift in the entire wage profile.

Without, then, carefully specifying the labor-supply model underlying the fixed-effect formulation, it is difficult to know exactly what combination of parameters is being estimated by the natural-experiment approach. Including variables in Q needed for an interpretation of γ invariably implies that one must rely on the generalized fixed-effect specification given by Eq. (5.4), meaning that covariates Z_{it} must be accounted for when calculating the difference-in-difference estimator. In addition to entering specifications directly, the presence of Z_{it} typically alters the formulation of $\Delta_t Q$ which further complicates the interpretation of γ.

Another critical qualification revealed by this attempt to interpret the difference-in-difference estimator involves the requirement that only the experimental group experiences the policy change. If controls also undergo a change at the same time, albeit a different change, then the appropriate specification for Eq. (5.1) becomes

$$y_{it} = \gamma_e \delta_{it} + \gamma_c(1 - \delta_{it}) + \eta_i + m_t + \varepsilon_{it}, \tag{5.11}$$

where γ_e and γ_c represent the behavioral response of the experimentals and controls, respectively.

Such a circumstance would arise, for example, in the case of the 1986 US tax reform. A particular change in the tax code may have directly impacted only a segment of taxpayers (experimentals), but many changes were made to the tax code simultaneously and literally all taxpayers were affected. This would also be the case if there were general equilibrium effects of the policy intervention that affected all wages (or prices) in the economy.

With the term $\gamma_c(1 - \delta_{it})$ present in Eq. (5.11), the difference-in-difference estimator $\hat{\gamma}$ loses its interpretation as a response to any policy change. Fixed-effect estimation of Eq. (5.11) directly can in principle, recover behavioral responses γ_e and γ_c with interpretations analogous to Eq. (5.10), but most formulations imply correlation between δ_{it} and ε_{it}, rendering least squares inconsistent. Such correlation arises when the size of the policy change is systematically different across experimentals and controls, and this occurs almost by definition since it is the nature of the policy change that distinguishes experi-

mentals and controls. With endogeneity induced by this correlation, instrumental-variable procedures must be implemented to estimate (5.11).

5.3. A review of some empirical applications

The empirical applications reviewed here all consider the impact of tax reforms on labor supply. The usual strategy adopted for estimation in these studies is to include the policy dummy with some controls for the wage, other income and demographic variables. As our analysis in Section 5.1 has shown, the interpretation of the estimates from these studies depends on which control variables were included and whether, for the groups chosen, the required assumptions on the unobservable error terms are plausible.

A difference-in-differences estimator was used by Eissa (1995a) to evaluate the effects of the US 1986 Tax Reform Act (TRA) on married women's labor supply. She uses the repeated cross sections of the March Current Population Surveys (CPSs) and compares data from the 1984–1986 surveys just preceding the reform and the 1990–1992 surveys sometime after. Her study compares the behavior of wives married to high-earning husbands (those who were at or above the 99th percentile of the CPS income distribution) to that of wives of lower earning husbands (between the 75th and 80th percentile of the income distribution). The two groups were affected differentially by the 1986 tax reform.

Estimates are provided for both participation and hours. In particular, a reduced form probit equation for participation and an hours equation which included an inverse Mills ratio control for selection were estimated. Demographic variables were entered in the model and some specifications allowed for interactions of the response coefficient with education level. These adjustments were found to significantly reduce the elasticity estimates. The reported wage elasticities for hours were between 0.6 and 1 while, for participation, elasticity estimates were surprisingly smaller – between 0.1 and 0.6. The choice of grouping is controversial since it might be thought that, even given the observed included controls, husband's income is not exogenous for the change in his spouse's labor supply. Moreover, given the increasing dispersion of incomes and wages among all groups during that period, the common time effects (common trends) assumption among the unobservable components across the two groups may not be satisfied.

Eissa's approach was also followed in a recent panel data study of the 1987 Danish tax reform by Graversen (1996). He considered the participation and hours worked of women, split according to marital status. In both cases, for controls he used a group for which predicted tax rate changes were small, using pre-reform hours and wages on the post-reform tax parameters. No exclusion restrictions appear to have been used to identify the selection term. For the difference-in-differences estimates with no controls for observable individual characteristics, he found perversely signed effects, but including numbers and age of children, for example, resulted in small but positive responses. This sensitivity of the difference-in-differences parameter estimates to the inclusion of observable time-varying characteristics is indicative of the importance of the conditions placed on the distribution of unobservables within each group over time.

Eissa and Liebman (1995) focused on the effects of TRA and EITC on single women with children. Again, they used data from the March CPSs for the US. Their identification strategy was to compare the change in labor supply for women with children to the change in labor supply for women with no children. They found that the participation of single women with children increased by 1.9–2.8 percentage points relative to single women with no children (from a base of 73%). Eissa and Lieberman also found a rather more surprising result that the EITC expansion in the Tax Reform Act had no perceptible effect on the hours of work of single women with children who were already in employment. The use of women with no children as a control group is open to criticism on a number of grounds. First, the conditions on the time and composition effects among the unobservables is unlikely to be satisfied in the repeated cross-sections of the CPS, even given the included regressors. Second, women with no children are probably working closer to their upper bound, as far as participation is concerned, and would not, therefore, be expected to increase participation. This is really a failure of the common trends assumption since such women may not, therefore, be able to absorb an upward common trend to labor supply on the participation margin.

Blundell et al. (1998b) consider the use of the sequence of tax reforms in the UK over the 1980s and early 1990s to study the hours responses of married women from a long time series of repeated cross-sections. A semi-log linear labor supply equation (see Eq. (4.54)) was specified with additive controls for other income, children, education and birth cohort. In contrast to the other studies discussed in this section, the hours equation included the log of the post-tax hourly wage rate and other income as well as a number of demographic controls. Other income was defined by the difference between consumption and the product of hours worked and the post-tax marginal hourly wage. This definition of other income is consistent both with intertemporal two-stage budgeting in the absence of liquidity constraints and with the presence of liquidity constraints as described in Section 4 above. The estimated labor supply model allowed the demographic variables to interact with the log wage and other income variables.

Two alternative estimators were considered. The first was a difference-in-differences estimator that grouped the sample by taxpayers and non-taxpayers. This was argued to be invalid because, under very general conditions, the composition of the two groups could be expected to change in a non-random way in response to the tax reforms. The second approach grouped by education and age cohort. This exploited the systematically changing distribution of wages by education and cohort group in the UK described in Section 3.4. The idea was that the differential growth in wages across birth cohorts by education group reflects changes in the demand for labor, possibly due to skill-biased technical change, and could be excluded from the labor supply equation. The log marginal hourly wage, which was included directly in the labor supply specification regression, together with the other income variable and participation in work, were treated as endogenous. The estimator can, thus, be interpreted as a (grouping) instrumental variable estimator in which the changes in the demand for the different skills of each education and cohort group are assumed to be exogenous and validly excluded from labor supply given the inclusion of the wage and

income variables. The education and cohort interactions with time, which were the excluded instruments, were found to be jointly significant in the wage and other income reduced forms.

The reported uncompensated labor supply elasticities, although small, were all positive and highest for women with children of pre-school age. The income elasticities were all negative, except for those women with no children, for whom they were essentially zero. As a result, the compensated wage effects, which matter for welfare, were all positive and the model was found to be consistent with standard theory everywhere in the data. In comparison, the estimates that use taxpayer status as a grouping instrument showed a significant *negative* wage elasticity. This negative estimate was argued to reflect the systematic change in the composition of the taxpaying group. During the period considered, there were many new entrants into this taxpaying group who had systematically lower hours. This non-random change in composition invalidates the second assumption of Section 5.1.2 on the composition of groups across time.

A number of additional experiments were reported that varied the control variables and instruments. In one experiment the time effects and the cohort/education effects were excluded. Only age and age squared were entered along with the demographic variables, the log marginal wage and the other income variable. This makes the specification similar to that in traditional cross section studies, such as those reviewed by Mroz (1987), except that the data contain a large number of time periods. Both the other income and wage elasticities became much larger. The resulting estimators are similar to those reported in the Arellano and Meghir (1992) study on the UK where education was used as the identifying instrument. As in the Eissa (1995b) study, controlling for education in the labor supply equation has the property of reducing the wage elasticity.

Eissa (1996) considers the case of labor supply responses to the sequence of tax reforms during the 1980s in the US. As in her previous studies, she uses March CPS data but this time over a longer period – 1976 to 1993. Moreover, the grouping was by education level. She finds only weak evidence of an increase in male labor supply in response to the Tax Reform Act. This poses an interesting issue relating to the larger effects on taxable income that have been found in studies that use tax-return data directly such as Feldstien (1995).[24]

6. Estimation with non-participation and non-linear budget constraints

To analyze how tax and welfare policies influence hours of work, there has been a steady expansion in the use of sophisticated statistical models characterizing distributions of discrete-continuous variables that jointly describe work and program participation. Considered at the forefront of research in this area, these models offer a natural mechan-

[24] This could be reconciled if it can be shown that certain groups of individuals respond to tax reforms on other margins.

ism for capturing the institutional features of both tax and welfare programs. This section describes how to estimate the effect of these programs on labor supply using these models.

These models build on standard approaches for dealing with censored and missing data. The basic principles underlying these approaches are well documented in the econometrics literature. In what follows, we provide explicit details on applying these methods to incorporate fixed costs, missing wages and discrete program participation in models of labor supply behavior with taxes and welfare.

6.1. Basic economic model with taxes

Consider a model of static labor supply where individuals determine hours of work and consumption by maximizing a utility function $U(C,h)$ subject to the budget constraint

$$C = Wh + Y - \tau(I), \tag{6.1}$$

where C is the consumption, W is the gross wage/h, h is the hours of market work, Y is the non-labor income, τ is the taxes determined by the function $\tau(\cdot)$, I is the taxable income per year, $I = Wh + Y - D$, and D is the deductions per year.

Due to different marginal tax rates in the various income brackets combined with the existence of non-labor income, the budget set is inherently non-linear in most instances. The literature applies two approaches for modeling the non-linearities induced by taxes: piecewise-linear functions that recount the brackets making up tax schedules; and smooth differentiable relations that summarize the tax rates implied by bracketed schedules. This section outlines each of these approaches, along with the procedures implemented to estimate labor-supply parameters associated with each approach.

In the absence of taxes, maximization of the utility function subject to the budget constraint defines the labor supply function

$$h = f(W, Y, v), \tag{6.2}$$

where v is an error reflecting the contribution of factors relevant to economic agents and unobserved by the econometrician.[25] With W and Y reinterpreted as "after-tax" measures, the construction of which is presented below, f continues to describe hours-of-work behavior even when complex non-linearities affect budget constraints, as is the case with taxes. The objective of most labor-supply analyses is to estimate the parameters of the function f.

6.1.1. Structure of taxes

The institutional features of income and program taxes occupy a great deal of attention in

[25] It is straightforward to replace Eq. (6.2) by $f(W,Y,X,v)$ where X is a vector incorporating measured variables affecting agents' choices. We suppress X for notational convenience. f, of course, depends on a parameter vector which we also suppress.

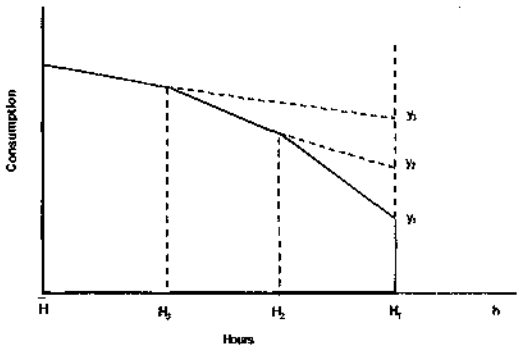

Fig. 15. Budget constraint with income taxes.

the labor supply literature. As described in Section 2, the complexities introduced by the US and the UK tax system, for example, contort the budget constraint faced by a typical worker. Modeling this constraint is often thought to be essential in labor supply analysis for capturing the opportunities available to individuals. For example, the overall tax schedule in the US consists of five components:

$$\tau(Y,E) = \text{FEDTX} + \text{STATX} + \text{EITC} + \text{SSTAX} + \text{WELFARE}, \tag{6.3}$$

where $\tau(Y,E)$ is the overall tax schedule, E is earned income, FEDTX is the federal income tax schedule, STATX is the state income tax schedule, EITC is the earned income tax credit schedule, SSTAX is the social security tax schedule, and WELFARE is net transfers from public assistance programs.

Each of these schedules has its own method of computing "taxable" income, but all in some way base calculations on a distinction between Y and E. We ignore these considera tions here. Both federal and state income tax schedules compute taxes based on income brackets, which induces piecewise linear budget constraints. The other programs are applicable over only part of the income range which also creates brackets.

6.1.2. Piecewise linear constraints

Fig. 15 shows a hypothetical budget constraint for an individual in the US faced with federal income taxes alone, state income taxes alone, or both.[26] In this diagram, h denotes hours of work, and "Consumption" denotes total after-tax income or the consumption of market goods. The segments of the budget constraint correspond to the different marginal tax rates that an individual faces. In particular, he faces a tax rate of t_A between H_0 hours and H_1 hours (segment 1) and tax rates of t_B and t_C, respectively, in the intervals (H_1, H_2) and $(H_2, \overline{H})$ (segments 2 and 3). Thus, the net wages associated with each segment are:

[26] Note that $h = H_0 = 0$ corresponds to 0 h of work. As we move from right to left in these figures, hours of work increase.

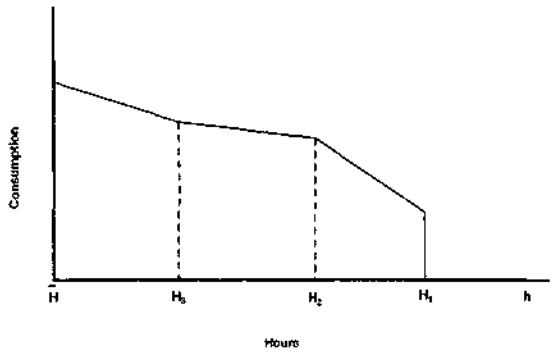

Fig. 16. Budget constraint with EITC.

$w_1 = (1 - t_A)W$ for segment 1, $w_2 = (1 - t_B)W$ for segment 2, and $w_3 = (1 - t_C)W$ for segment 3. Virtual income for each segment (i.e., income associated with a linear extrapolation of the budget constraint) is calculated as: $y_1 = Y - \pi(Y, 0)$; $y_2 = y_1 + (w_1 - w_2)H_1$; and $y_3 = y_3 + (w_2 - w_3)H_2$. Changes in tax brackets create the kink points.

Fig. 16 shows a budget constraint affected only by the EITC schedule,[27] and Fig. 17 shows a budget constraint that reflects the effects of the social security tax alone.[28] As seen in Fig. 18, welfare benefit programs create a budget set that resembles the one for Social Security. All of these taxes induce non-convexities in opportunity sets.

Summing these various tax components creates an overall tax-transfer schedule with two noteworthy features. First, the schedule faced by a typical individual includes a large number of different rates. Translated into the hours-consumption space, this implies a large number of kink points in the budget constraint. Second, for most individuals the tax schedule contains non-convex portions, which arise from four potential sources. The first arises from a fall in the EITC tax rate at the break even point. In Fig. 16, that point occurs at H_2 where the tax rate falls from a positive value to zero. The second source occurs when the social security tax hits its maximum (at H_1 in Fig. 17), where the corresponding tax rate goes from a positive value to zero. A third source is the non-convexity introduced by the structure of the standard deduction. Finally, if a worker's family participates in any welfare program, then significant non-convexities arise as benefits are withdrawn when earnings increase.

[27] The EITC is a negative income tax scheme which can induce, in the simplest case, two kinks in a person's constraint: one where the proportional credit reaches its maximum (H_1 in Fig. 16), and one at the break even point where the credit is fully taxed away (H_2 in the figure). The tax rates associated with the first two segments are t_A, which is negative, and t_B, which is positive. Thereafter, the EITC imposes no further tax.

[28] The social security tax is a proportional tax on earnings up to a specified earnings level, after which the amount of tax paid is the same regardless of earnings. As a result, Fig. 17 shows a constraint with a single interior kink (given by H_1 in the figure) corresponding to the maximum proportionally taxed earnings level. The tax rate on the segment leading up to that kink is t_A, switching to zero on the second segment.

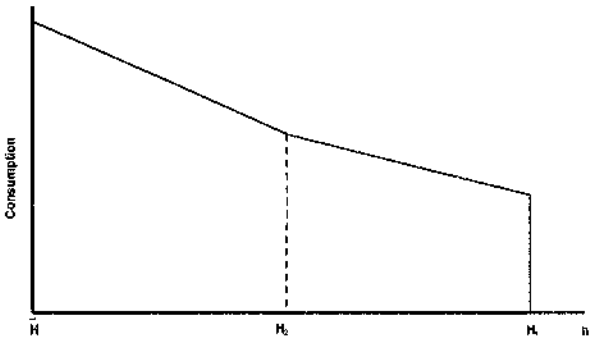

Fig. 17. Budget constraint with Social Security Tax.

6.1.3. Constructing a differentiable constraint

Approximating the tax schedule by a differentiable function leads to a simple approach for developing an empirical model of labor supply that recognizes the influence of taxes. A convenient approach for constructing this function is to approximate the marginal tax rate schedule – a step function – by a differentiable function. This approximation must itself be easily integrable to obtain a simple closed form for the tax function.

An elementary candidate for constructing a differentiable approximation that can be made as close as one desires to the piecewise-linear tax schedule has been applied in MaCurdy et al. (1990). To understand the nature of the approximation, return to Fig. 15. One can represent the underlying schedule as follows:

$$\begin{aligned} \tau'(I(h)) &= t_A \qquad \text{from } I(H_0) \text{ to } I(H_1) \\ &= t_B \qquad \text{from } I(H_1) \text{ to } I(H_2) \\ &= t_C \qquad \text{above } I(H_2), \end{aligned} \tag{6.4}$$

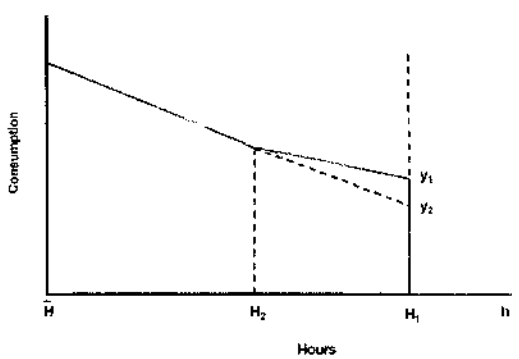

Fig. 18. Budget constraint with welfare.

where $\tau'(I(h))$ is the marginal tax rate, $I(h)$ is taxable income at h hours of work, and t_i is the marginal tax rate, $i = A, B, C$.For expositional simplicity, suppose that $t_A = 0$. Consider the following approximation of this schedule which uses three flat lines at the heights $t_A (= 0)$, t_B and t_C and weight functions parameterized to switch the three lines on and off at appropriate points:

$$\hat{\tau}'(I(h)) = t_B[\Phi_1(I(h)) - \Phi_2(I(h))] + t_C[\Phi_2(I(h))], \tag{6.5}$$

where the weight functions are given by $\Phi_i(I(h)) =$ the cumulative distribution function with mean μ_i and variance σ_i^2, $i = 1, 2$. The middle segment of the tax schedule has height t_B and runs from taxable income $I(H_1)$ to $I(H_2)$. To capture this feature, parameterize $\Phi_1(\cdot)$ and $\Phi_2(\cdot)$ with means $\mu_1 = I(H_1)$ and $\mu_2 = I(H_2)$, respectively, with both variances set small. The first distribution function, $\Phi_1(\cdot)$ takes a value close to zero for taxable income levels below $I(H_1)$ and then switches quickly to take a value of one for higher values. Similarly, $\Phi_2(\cdot)$ takes a value of zero until near $I(H_2)$ and one thereafter. The difference between the two equals zero until $I(H_1)$, one from $I(H_1)$ to $I(H_2)$ and zero thereafter. Thus, the difference takes a value of one just over the range where t_B is relevant. Notice that we can control when that value of one begins and ends by adjusting the values μ_1 and μ_2. Also, we can control how quickly this branch of the estimated schedule turns on and off by adjusting the variances of the cumulative distribution functions, trading off a more gradual, smoother transition against more precision. In general, adjusting the mean and variance parameters allows one to fit each segment of a schedule virtually exactly, switch quickly between segments, and still maintain differentiability at the switch points.

A generalization of this approximation takes the form

$$\hat{\tau}'(I(h)) = \sum_{i=1}^{k} [\Phi_i(I(h)) - \Phi_{i+1}(I(h))]b_i(I(h)), \tag{6.6}$$

where the functions $b_i(I(h))$ are polynomials in income. With the Φ_i denoting normal c.d.f.s, function (6.6) yields closed form solutions when it is either integrated or differentiated.[29] The resulting approximation can be made to look arbitrarily close to the budget constraint drawn in Fig. 15, except that the kink points are rounded.

6.2. *Instrumental-variable estimation*

Conventional non-linear instrumental-variable procedures offer a robust method for estimating particular forms of the labor-supply function f in Eq. (6.2), forms that permit the specification of structural equations that are linear in all sources of disturbances. As discussed in Section 4, the development of such specifications is a substantial challenge

[29] Total taxes are given by: $\tau(I) = \int \tau'(I)\mathrm{d}I$. The following relations enable one to calculate an explicit form for $\tau(X)$: $\int \Phi \mathrm{d}I = I\Phi + \varphi$, $\int I\Phi \mathrm{d}I = (1/2)I^2\Phi - (1/2)\Phi + (1/2)I\varphi$, $\int I^2\Phi \mathrm{d}I = (1/3)I^3\Phi - (2/3)\varphi + (1/3)I^2\varphi$, $\int I^3\Phi \mathrm{d}I = (1/4)I^4\Phi - (3/4)\Phi + (3/4)I\Phi + (1/4)I^3\varphi$. In this expression, Φ refers to any Φ_i's, and φ designates the density function associated with Φ_i.

for it proves difficult to discover a preference map that produces additivity in structural disturbances – errors reflecting unobserved differences among people (heterogeneity) – while at the same time permitting measurement errors in hours and wages to enter linearly.

6.2.1. A useful characterization of labor supply with taxes

The introduction of a non-linear tax schedule into a model of labor supply poses few analytical difficulties when the schedule generates a strictly convex constraint set with a twice-differentiable boundary. Utility maximization in this case implies a simple characterization of the hours-of-work choice.

With τ denoting the smooth function that approximates the tax schedule, specify the marginal wage rate and "virtual" income as

$$\omega = \omega(h) = (1 - \tau')W,$$

$$y = y(h) = Y + E - \tau - \omega h = Y + \tau' Wh - \tau = C - \omega h, \tag{6.7}$$

where $E = Wh$ is gross earnings, and τ and τ' (the derivative of the tax function with respect to income) are evaluated at income level $I = I(h) = Y + Wh - D$ which directly depends on the value of h. In Eq. (6.7) we write the marginal wage $\omega = \omega(h)$ and virtual income $y = y(h)$ as functions to emphasize their dependence on hours h.

Utility maximization implies a solution for hours of work that obeys the implicit equation

$$h = f(\omega(h), y(h), v), \tag{6.8}$$

where we write the marginal wage $\omega = \omega(h)$ and virtual income $y = y(h)$ as functions to emphasize their dependence on hours. Figs. 15 and 18 illustrate this representation of the solution for optimal hours of work. This characterization follows from work on taxes and labor supply (e.g., Hall, 1973) that represents a consumer as facing a linear budget constraint in the presence of non-linear tax programs. This linear constraint is constructed in a way to make it tangent to the actual non-linear opportunity set at the optimal solution for hours of work. The implied slope of this linearized constraint is $\omega(h)$ and the corresponding value of virtual income is $y(h)$. Eq. (6.8) constitutes a structural relationship that determines hours of work. By applying the Implicit Function Theorem to specification (6.8), we can solve this implicit equation for h in terms of W, Y, and other variables and parameters entering the functions τ and f. This operation produces the labor supply function applicable in the non-linear tax case.

6.2.2. A structural equation of labor supply with taxes

Relation (6.8) directly provides the basis for formulating a structural equation that can be estimated by standard instrumental-variable procedures. Consider, for example, the semi-log specification[30]:

[30] See Eq. (A.5) in Appendix A.

$$h = f(\omega, y, v) = \mu + Z\gamma + \alpha \ln\omega + \beta y + v, \tag{6.9}$$

where μ, γ, α, and β are parameters, Z is a vector of observed determinants of labor supply (e.g., age, family size, etc.), and v is a structural disturbance capturing unobserved factors influencing hours-of-work decisions. The marginal after-tax wage ω enters this specification in a natural log, so α represents a hybrid of an uncompensated substitution effect and elasticity. The coefficient β corresponds to an income effect.

Conventional instrumental-variable procedures offer a robust method for estimating the coefficients of the semi-logarithmic specification of the labor supply function given by Eq. (6.9). In the absence of measurement error, inspection of Eq. (6.9) reveals that the error term enters linearly into the specification. Consequently, variables that are orthogonal to the structural disturbance v can serve as instruments for estimating the parameters determining substitution and income effects. The implementation of such procedures imposes no parametric restrictions and it allows one to consider a wide variety of exogeneity assumptions.

In many data sets there are serious suspicions that hours of work and wages are reported with error. Suppose h^* denotes measured hours of work and that the function $h^*(h,\varepsilon)$ relates h^* to actual hours, h, and to an error component, ε. An interesting specification for characterizing the form of reporting error is given by the multiplicative structure:

$$h^* = h^*(h, \varepsilon) = he^{\varepsilon}, \qquad \text{with } W^* = E/h^*, \tag{6.10}$$

where ε is distributed independently of h and v, and the distribution of the measurement error component ε satisfies the moment condition $E(e^{\varepsilon}) = 1$, implying that h^* and h have the same expected value. Whereas $W = E/h$ defines the true hourly wage rate, W^* designates the data available on wages. This formulation presumes not only measurement error in hours, but also the existence of reporting error in hourly wage rates due to its construction. Assuming accurate observations on E, calculating wages by dividing total labor earnings by reported hours induces a reciprocal relation in the measurement error linking data on hours and wages.

Incorporating the multiplicative measurement error model given by Eq. (6.10) into the semi-logarithmic specification of labor supply presented in Eq. (6.9) yields the empirical relations

$$h^* = \tilde{u} + Z\gamma + \alpha \ln\omega^* + \beta y + u, \tag{6.11}$$

where

$$\ln\omega^* = \ln(E/h^*) + \ln(1 - \tau^r),$$

$$\tilde{\mu} = \mu - \alpha\sigma_\varepsilon^2/2, \tag{6.12}$$

$$u = v + \alpha(\varepsilon + \sigma_\varepsilon^2/2) + (h^* - h) = v + \alpha(\varepsilon + \sigma_\varepsilon^2/2) + h(e^{\varepsilon} - 1).$$

Relation (6.7) continues to define the variable y. This virtual income quantity and the

marginal tax rate τ^{l} are not contaminated by measurement error because they are functions of Y, E and τ^{l} which are known without errors. The variable $\ln\omega^{*}$ represents the natural logarithm of the after-tax wage rate evaluated at observed hours, which differs from the actual marginal wage due to the presence of reporting error in hours. The disturbance u possesses a zero mean since $E(\varepsilon) = -\sigma_{\varepsilon}^{2}/2$, $E(e^{\varepsilon}) = 1$, and the error ε is distributed independently of all endogenous components determining h, including the heterogeneity disturbance v.

Interpreting relation (6.11) as a structural equation describing labor supply, instrumental-variable methods continue to offer a flexible scheme for consistent estimation of substitution and income parameters. Due to the heteroscedasticity of the disturbance u in Eq. (6.11), the estimation procedure must compute robust standard errors to produce valid test statistics. For consistent estimation of the parameters of Eq. (6.11), one needs to be able to identify a set of variables X that are orthogonal to the structural disturbance v, independent of measurement error ε, and are capable of predicting the endogenous variables ω^{*} and y. Selecting alternative formulations of X offers the opportunity to entertain a variety of exogeneity assumptions, even with measurement error present, thereby indicating the direction of potential biases in estimated work disincentive effects arising from these assumptions. The maximum-likelihood approaches discussed below typically maintain that all sources of income are exogenous determinants of work hours, including W, and Y (i.e., the gross wage rate, non-taxable non-labor income, and non-labor taxable income). Judicious inclusion and exclusion of these income sources in X provides a basis for judging whether endogeneity of wages and/or incomes is a problem. Of course, the ability to test these exogeneity assumptions critically relies on there existing a sufficient number of elements in X that satisfy exclusion restrictions in Eq. (6.11).

6.2.3. *Lifecycle considerations*

As outlined in Section 4, substituting an alternative measure for the variable y in Eq. (6.9) creates a labor-supply specification that is consistent with decision-making in a lifecycle context, and this specification can in turn be modified to account for the existence of income taxes.[31] In a static analysis with taxes, one specifies virtual income as

$$y = Q - \omega h, \qquad \text{with } Q = Y + E - \tau, \tag{6.13}$$

where the income components making up the quantity Q represent current income in the period. In such an analysis $Q = C$ by assumption. However, in an intertemporal setting it need not be the case that $y = Y + E - \tau - \omega h$ since one can have $C - \omega h \neq Y + E - \tau - \omega h$ due to saving or borrowing. As shown in Section 4, in a multiperiod framework, with or without an uncertain future, the construction of the virtual income variable y (or, more precisely, the quantity Q) must account for net savings in the period.

Given the availability of data for each family's total consumption, a formulation for Q that obviously accomplishes this task is to set

[31] For details beyond the discussion presented in this subsection, see MaCurdy (1983).

$$y = Q - \omega h, \qquad \text{with } Q = C \tag{6.14}$$

as the measure of virtual income. Given this construction for y, and lifetime utility maximization with strongly separable preferences over time, the function $h = f(\omega, yv)$ characterizes the optimal lifecycle choice of labor supply in the period under consideration.

Standard two-stage least squares procedures continue to provide a computationally simple method for consistently estimating the parameters of the function f, assuming, of course, that the empirical specification of f is linear in disturbances–such as specification (6.9). One can apply linear or non-linear instrumental-variable procedures to estimate coefficients depending on whether the specification of f is linear or non-linear in parameters, with robust standard errors computed when appropriate.

6.3. Maximum likelihood: convex differential constraints with full participation

Maximum-likelihood estimation of labor-supply models with a tax schedule described by a twice-differentiable boundary implying a convex budget set poses few difficulties. Provided the gross wage variable and the other income variable are assumed free from measurement error and independent of unobserved heterogeneity, such an estimation approach need not heavily rely on exclusion restrictions to identify parameters. In contrast to the case when implementing instrumental-variable procedures, even though marginal wages and virtual incomes are endogenous, non-linearities introduced through distributional assumptions provide a valuable source of identification. Because exclusion restrictions are often difficult to justify, many researchers turn to maximum likelihood to avoid making ad hoc exclusion properties. Of course, the independence assumptions on the distribution of unobserved heterogeneity in these maximum likelihood approaches are strong and are precisely what is being relaxed in the fixed effects models that underlie the difference-in-differences and related approaches outlined in Section 5.

6.3.1. Specification of likelihood functions with multiplicative measurement error

Considering maximum-likelihood estimation of the model analyzed in Section 6.2, suppose the heterogeneity-error-component v in the labor-supply function (6.9) and the disturbance ε in the measurement-error equation (6.10) for hours of work is independent of the gross wage and other income and possesses the joint distribution: $(v, \varepsilon) \approx g_{v\varepsilon}$, where $g_{v\varepsilon}$ designates a density function. Using relations (6.9) and (6.10) to perform a standard change in variables from the errors v and ε to the variables h and h^* produces the likelihood function needed to compute maximum-likelihood estimates. The transformation from (v,ε) to (h,h^*) is monotonic for a wide range of functional forms for f as long as the underlying preferences satisfy quasiconcavity and budget sets are convex.

Without measurement error, the likelihood function for hours of work, h, takes the form

$$l = \frac{\mathrm{d}v}{\mathrm{d}h} g_v(h - \mu - Z\gamma - \alpha \ln W - \alpha \ln(1 - \tau^l) - \beta y), \tag{6.15}$$

where g_v is the marginal density for v, and the Jacobian term is

$$\frac{\mathrm{d}v}{\mathrm{d}h} = 1 + \left(\left(\frac{\alpha}{W(1-\tau')} - \beta h \right) W^2 \frac{\partial \tau'}{\partial I} \right), \tag{6.16}$$

which is required to be non-negative. In Eqs. (6.15) and (6.16), the derivative τ' is evaluated at $I = Wh + Y - \tau(Wh + Y)$.

With multiplicative measurement error, the likelihood function for observed hours h^* becomes

$$l = \int_0^{\text{maxwage}} \int_0^{\text{maxhours}} \frac{\mathrm{d}v}{\mathrm{d}h} g_{v\varepsilon}(\ln h^* - \ln h, h - \mu - Z\gamma - \alpha \ln \omega - \beta y) \psi(W) \mathrm{d}h \mathrm{d}W, \tag{6.17}$$

where integration occurs over the hourly wage, which is unobserved, using its density $\psi(W)$.[32,33] The non-negativity of the Jacobian term clearly places restrictions on the behavioral parameters and we discuss these restrictions further below.

6.3.2. A popular linear empirical specification

One of the most numerous commonly applied empirical specification for labor supply implemented in maximum likelihood analyses – particularly those using the piecewise-linear approach discussed below – takes the linear form:

$$h = f(\omega, y, v) = \mu + \alpha\omega + \beta y + Z\gamma + v \equiv \hat{h} + v, \tag{6.18}$$

where the unobserved error component v represents heterogeneity in preferences with $v \approx g_{\&\mathrm{nu};}$, where g_v denotes the marginal density of v. In conjunction with this specification, analyses also presume measurement error in hours of work possessing the classical

[32] This likelihood function fundamentally differs from the one proposed in Eq. (D.5) in Appendix D of MaCurdy et al. (1990). The particular form of the labor-supply model considered in MaCurdy et al. (1990) is

(a) $h^* = \tilde{\mu} + Z\gamma + \alpha\omega^* + \beta y + u$

with $h^* = he^{\varepsilon}$. The analog of (6.17) for this linear specification is

(b) $l = \int_0^{\text{maxwage}} \int_0^{\text{maxhours}} \frac{\mathrm{d}v}{\mathrm{d}h} (\ln h^* - \ln h, h - \mu - Z\gamma - \alpha\omega - \beta y) \psi(W) \mathrm{d}h \mathrm{d}W$

with $(\mathrm{d}v/\mathrm{d}h) = 1 + (\alpha - \beta h)W^2(\partial\tau'/\partial I)$. Likelihood function (b) is the valid specification for estimating model (a), whereas likelihood function (D.5) presented in MaCurdy et al. is not – unbeknownst, unfortunately, to the authors of MaCurdy et al. Specification (D.5) of MaCurdy et al. implicitly conditions on the true wage rate W, even though earnings, E, rather than W appears in (D.5). W is an unobserved variable in the analysis and, therefore, must be integrated out of (D.5) to obtain a valid formulation. Specifications (b) and (6.17) incorporate this integration. MaCurdy recognized this oversight when reconciling some Monte Carlo findings done by Lennart Flood during his visit to Stanford in 1996; Lennart's assistance in revealing this problem is gratefully acknowledged.

[33] If W is not independent of v and ε, then (6.17) is replaced by

$$l = \int_0^{\text{maxwage}} \int_0^{\text{maxhours}} \frac{\mathrm{d}v}{\mathrm{d}h} g_{v\varepsilon W}(\ln h^* - \ln h, h - \mu - Z\gamma - \alpha \ln \omega - \beta y, W) \mathrm{d}h \mathrm{d}W,$$

where $g_{v\varepsilon W}$ is the joint density of v, ε, and W.

linear functional form

$$h^* = h^*(h, \varepsilon) = h + \varepsilon, \tag{6.19}$$

where $\varepsilon \sim g_\varepsilon$, with ε and v independent. The measurement error component ε represents reporting error that contaminates observations on h for individuals who work.

The derivation of likelihood functions for this case is straightforward given the assumptions about preferences and budget constraints maintained to this point. Assuming no measurement error (i.e., $h^* = h$), a change in variables from the heterogeneity error v to actual hours h using relation (6.18) yields the likelihood function for h:

$$g^h(h) = \frac{\mathrm{d}v}{\mathrm{d}h} g_v(h - \mu_v - Z\gamma - \alpha\omega - \beta y), \tag{6.20}$$

where the Jacobian term is

$$\frac{\mathrm{d}v}{\mathrm{d}h} = 1 + (\alpha - \beta h)W^2 \frac{\partial \tau'}{\partial I}. \tag{6.21}$$

This Jacobian term is restricted to be non-negative over the admissible range. Maximizing (6.20) yields maximum-likelihood estimates for the parameters of the labor supply function f, which provide the information needed to infer the work disincentive effects of taxation.

If hours are indeed contaminated by additive measurement error, then the likelihood function for observed hours $h^* = h + \varepsilon$ is given by:

$$g_{h^*}(h^*) = \int_0^{\text{maxhours}} g_\varepsilon(h^* - h) g_h(h) \mathrm{d}h. \tag{6.22}$$

This expression resembles relation (6.20) except that integration occurs over hours to account for the existence of reporting error, and h^* replaces actual hours h in the Jacobian term in (6.17).

6.3.3. Imposition of behavioral restrictions with differentiable constraints

The implementation of maximum likelihood procedures imposes interesting and important restrictions on behavioral parameters in the presence of non-linear budget constraints. Consider, for example, likelihood function (6.22). For this specification to be a properly-defined likelihood functions, the Jacobian (6.21) must be non-negative. Violation of this condition implies that the density function for h is negative, which obviously cannot occur. Relation (6.20) indicates that this non-negativity condition translates into the property

$$\frac{\partial h^s}{\partial w} - \frac{\partial h^s}{\partial y} h \geq -\left(\frac{\partial \tau}{\partial I} W^2\right)^{\mathrm{I}} \leq 0, \tag{6.23}$$

where h^s $(=f)$ refers to the labor supply function. The left-hand side of this inequality is the Slutsky term. This inequality result does not require compensated substitution effects to be

positive as quasi-concave preferences mandate, only that these effects cannot become too negative.

Maximum likelihood procedures yield nonsensical results unless Eq. (6.23) holds. Without measurement error, estimated parameter values cannot imply a violation of Eq. (6.23) at any of the data combinations $(h,w(h),y(h))$ actually observed in the sample. If a violation occurs, then the evaluation of Eq. (6.22) for the observation associated with this combination would result in a non-positive value which causes the overall log likelihood function to approach minus infinity-which clearly cannot represent a maximum. With measurement error, maximum likelihood estimation applied to Eq. (6.22) ensures that a weighted average of Eq. (6.22) holds, with weighting occurring over all combinations of hours, marginal wages, and virtual income lying in the feasible range of the budget constraint of any individual included in the sample. Since maximum likelihood procedures assume the validity of such restrictions when calculating estimates of the coefficients of h^s, the resulting estimated labor supply function can be expected to exhibit compensated substitution effects that obey inequality (6.23) over a very wide range of hours, wages, and incomes.[34] Section 6.4.3 revisits these restrictions, relating them to those invoked in cases when maximum likelihood is used with non-differentiable (piecewise-linear) tax functions.

6.4. Maximum likelihood: convex piecewise-linear constraints with full participation

The majority of empirical labor-supply studies incorporating taxes treat the tax schedule as a series of brackets implying a piecewise-linear budget set. With such a tax function, the familiar change-in-variables techniques implemented in conventional maximum likelihood do not apply due to the non-existence of the Jacobian over measurable segments of the sample space, which occurs since the functional relationships characterizing hours-of-work choices are not differentiable. Moreover, a piecewise-linear budget set creates endogenous variables (hours and after-tax wages) that are both discrete and continuous in character, complicating the use of instrumental-variable procedures, which require the inclusion of sample-selection terms in equations to produce disturbances with zero means.

6.4.1. Characterization of labor supply with piecewise-linear constraints

To illustrate the derivation of an estimable labor supply model using the piecewise-linear approach for the model described in Section 6.1.2., consider the simple case of a budget set with only three segments as presented in Fig. 15. The preceding discussion defines the variables y_j, ω_j, and H_j appearing in this figure. To locate the kinks and slopes of the budget constraint for an individual, a researcher must know the individual's level of non-labor

[34] It is, of course, computationally feasible to use (6.22) in estimation and not require g_h to be defined over the entire range of its support. Computationally one merely requires g_h to be non-negative over a sufficiently large region to ensure (6.22) $>$ 0. Of course, not requiring $g_h \geq 0$ over its relevant range produces a nonsensical statistical model.

income, gross wage rate, hours of work, and the structure of the tax system. The hours of work at which kinks occur are given by $H_j = (I_j - Y + D)/W$, where Y and D, respectively, represent taxable non-labor income and deductions, and I_j is the maximum taxable income for segment j. The slope of each segment is given by the marginal wage rate for that segment: $\omega_j = W(1 - t_j)$, where j denotes the segment, t_j signifies the marginal tax rate for that segment, and W is the gross wage rate/h. Finally, the non-labor income at zero hours of work – the intercept of the budget line – is $y_1 = V + Y - \tau(Y - D)$, where $\tau(\cdot)$ is the tax function evaluated at the individual's taxable income at zero earnings. Given this intercept value, virtual incomes or the intercepts associated with successive budget segments are computed by repeated application of the formula: $y_j = y_{j-1} + (\omega_{j-1} - \omega_j)H_{j-1}$.

Given a convex budget constraint, an individual's optimization problem amounts to maximizing $U(C,h)$ subject to

$$\begin{aligned}
C &= y_1 && \text{if } h = 0,\\
&= \omega_1 h + y_1 && \text{if } H_0 < h \leq H_1,\\
&= \omega_2 h + y_2 && \text{if } H_1 < h \leq H_2,\\
&= \omega_3 h + y_3 && \text{if } H_2 < h \leq \bar{H},\\
&= \omega_3 \bar{H} + y_3 && \text{if } h = \bar{H},
\end{aligned} \tag{6.24}$$

The solution of this maximization problem decomposes into two steps. First, determine the choice of h conditional on locating on a particular segment or a kink. This step yields the solution

$$\begin{aligned}
h &= 0 && \text{if } h = 0 && \text{(lower limit)},\\
&= f(\omega_1, y_1, v) && \text{if } 0 < h < H_1 && \text{(segment 1)},\\
&= H_1 && \text{if } \quad h = H_1 && \text{(kink 1)},\\
&= f(\omega_2, y_2, v) && \text{if } H_1 < h < H_2 && \text{(segment 2)},\\
&= H_2 && \text{if } h = H_2 && \text{(kink 2)},\\
&= f(\omega_3, y_3, v) && \text{if } H_2 < h < \bar{H} && \text{(segment 3)},\\
&= \bar{H} && \text{if } h = \bar{H} && \text{(kink 3 = upper limit)}.
\end{aligned} \tag{6.25}$$

Second, determine the segment or the kink on which the person locates. The following relations characterize this solution: choose

$$
\begin{aligned}
&\text{if } f(\omega_1, y_1, v) \leq 0 && 0,\\
&\text{if } H_0 < f(\omega_1, y_1, v) < H_1 && \text{(Segment 1)},\\
&\text{if } f(\omega_2, y_2, v) \leq H_1 < f(\omega_1, y_1, v), && \text{(Kink 1)},\\
&\text{if } H_1 < f(\omega_2, y_2, v) < H_2 && \text{(Segment 2)},\\
&\text{if } f(\omega_3, y_3, v) \leq H_2 < f(\omega_2, y_2, v) && \text{(Kink 2)}\\
&\text{if } H_2 < f(\omega_3, y_3, v) < \bar{H} && \text{(Segment 3)},\\
&\text{if } f(\omega_3, y_3, v) \geq \bar{H} && \text{(Kink 3)}.
\end{aligned} \tag{6.26}
$$

Combined, these two steps imply the values of h and C that represent the utility-maximizing solutions for labor supply and consumption.

6.4.2. Specification of the likelihood function with measurement error: all participants
The linear specification of f given by Eq. (6.18) implies the following stochastic specification for labor supply:

$$
\begin{aligned}
&\hat{h}_1 + v + \varepsilon && \text{if } 0 < \hat{h}_1 + v \leq H_1 && \text{(segment 1)},\\
&H_1 + \varepsilon && \text{if } \hat{h}_2 + v < H_1 < \hat{h}_1 + v && \text{(kink 1)},\\
&h^* = \hat{h}_2 + v + \varepsilon && \text{if } H_1 < \hat{h}_2 + v \leq H_2 && \text{(segment 2)},\\
&H_2 + \varepsilon && \text{if } \hat{h}_3 + v < H_2 < \hat{h}_2 + v && \text{(kink 2)},\\
&\hat{h}_3 + v + \varepsilon && \text{if } H_2 < \hat{h}_3 + v \leq \bar{H} && \text{(segment 3)},\\
&\bar{H} + \varepsilon && \text{if } \hat{h}_3 + v \geq \bar{H} && \text{(upper limit)}.
\end{aligned} \tag{6.27}
$$

This represents a sophisticated variant of an econometric model that combines discrete and continuous choice elements.

All studies implementing the piecewise-linear approach assume the existence of measurement error in hours of work. With the linear measurement error model given by Eq. (6.19), observed hours $h^* = h + \varepsilon$. As long as the measurement error component ε is continuously distributed, so is h^*. In contrast to information on h, knowledge of h^* suffices neither to allocate individuals to the correct branches of the budget constraints nor to identify the marginal tax rate faced by individuals, other than at zero hours of work. The state of the world an individual occupies can no longer be directly observed, and one confronts a discrete data version of an errors-in-variables problem. The interpretation of

measurement error maintained in this analysis is that ε represents reporting error that contaminates the observation on h for persons who work.[35]

The log-likelihood function for this model is given by $\Sigma_i \log g_{h^*}(h_i^*)$, where i indexes observations. Defining $v_j = H_{j-1,i} - \hat{h}_{ji}$ and $\bar{v}_j = H_{j,i} - \hat{h}_{ji}$, the components $g_{h^*}(h_i^*)$ are given by

$$g_{h^*}(h^*) = \sum_{j=1}^{3} \int_{\underline{v}_j}^{\bar{v}_j} g_2[h^* - \hat{h}_j, v]\mathrm{d}v \qquad \text{(segments 1, 2, 3)},$$

$$+ \sum_{j=1}^{2} \int_{\bar{v}_j}^{\underline{v}_{j+1}} g_1[h^* - H_j, v]\mathrm{d}v \qquad \text{(kinks 1, 2)},$$

$$+ \int_{\bar{v}_3}^{\infty} g_1[h^* - \bar{H}, v]\mathrm{d}v \qquad \text{(upper limit)}, \tag{6.28}$$

where $g_1(\cdot,\cdot)$ and $g_2(\cdot,\cdot)$ are the bivariate density functions of (ε,v) and $(\varepsilon + v,v)$, respectively. Maximizing the log-likelihood function produces estimates of the coefficients of the labor supply function f. These estimates provide the information used to infer both substitution and income responses, which in turn provide the basis for calculating the work disincentive effects of income taxation.

6.4.3. Comparisons of the piecewise-linear approach with other estimation procedures

The piecewise-linear approach for estimating the work disincentive effects of taxes offers both advantages and disadvantages relative to other methods. Concerning the attractive features of this approach, piecewise-linear analyses recognize that institutional features of tax systems induce budget sets with linear segments and kinks. This is important if one believes that a smooth tax function does not provide a reasonably accurate description of the tax schedule. The piecewise-linear approach admits randomness in hours of work arising from both measurement error and variation in individual preferences and it explicitly accounts for endogeneity of the marginal tax rate in estimation, but so do the instrumental-variable and differentiable likelihood methods discussed above. As we will see below, the piecewise-linear approach more readily incorporates fixed costs of holding a job, regressive features of the tax code, and multiple program participation than other procedures due to the discrete-continuous character of hours-of-work choices induced in these environments. These features of the piecewise-linear method make it a vital approach in empirical analysis of labor supply.

[35] Note that expected hours of work, in this convex piece-wise linear case, is additive in each hours choice weighted by the probability of each segment or kink. Each term in this sum being at most a function of two marginal wages and two virtual incomes. Blomquist and Newey (1997) exploit this observation to develop a semi-parametric estimator for hours of work with piece-wise linear taxation, imposing the additivity through a series estimator.

On the other hand, the following shortcomings of the piecewise-linear procedure raise serious doubts about the reliability of its estimates of work disincentive effects. First, the piecewise-linear methodology assumes that both the econometrician and each individual in the sample have perfect knowledge of the *entire* budget constraint that is relevant for the worker in question. Errors are permitted neither in perceptions nor in measuring budget constraints. Taken literally, this means that: all income and wage variables used to compute each sample member's taxes are observed perfectly by the econometrician; individuals making labor supply choices know these variables exactly prior to deciding on hours of work; each individual and the econometrician know when the taxpayer will itemize deductions and the amount of these itemizations; and each taxpayer's understanding of the tax system is equivalent to that of the econometrician (e.g., the operation of such features as earned-income credits). Clearly, given virtual certainty that most of these assumptions are violated in empirical analyses of labor supply, the estimates produced by methods relying on these assumptions must be interpreted very cautiously. The differentiable-likelihood methods rely on the same assumptions. The instrumental-variable methods do not, so they are likely to be more robust.

Second, measurement error plays an artificial role in econometric models based on the piecewise-linear approach. Its presence is needed to avoid implausible predictions of the model. The statistical framework induced by the piecewise-linear approach implies that bunching in hours of work should occur at kink points if hours precisely measure h. However, for the vast majority of data sources currently used in the literature, only a trivial number of individuals, if indeed any at all, report hours of work at interior kink points. Unless one presumes that the data on hours do not directly represent h, such evidence provides the basis for immediately rejecting the distributional implications of the above specifications. Considering, for example, the labor-supply characterization proposed in Eq. (6.27), almost any test of the distributional assumptions implied by this specification would be readily rejected because observed hours would take the values H_0, H_1, H_2, and $\underline{H}$ with only a trivial or zero probability. Instead, observed hours essentially look as if they are distributed according to a continuous distribution. When a continuously-distributed measurement error ε is added to the model, observed hours h^* are continuously distributed. This provides an essential reason for introducing measurement error in the data, for without it the piecewise-linear structure provides a framework that is grossly inconsistent with the data. Of course, several sound reasons exist for admitting measurement error in a labor supply model, including the widespread suspicion that reporting error contaminates data on hours of work. However, measurement error in hours of work implies measurement error in wages, since they are typically computed as average hourly earnings. Current applications of the piecewise-linear analysis mistakenly ignore this by assuming perfectly measured budget constraints.[36] The unnatural role played by measurement error raises questions about the credibility of findings derived from the piecewise-

[36] It is possible to argue that this error does not result in measurement error in the hourly wage, if the measurement error is interpreted as an "optimization" error.

linear approach.[37] In contrast to the piecewise-linear approach, it is not essential to introduce measurement error in either the differentiable-likelihood or the instrumental-variable approach because hours in the distribution of h are continuous without measurement error.

Third, existing research implementing the piecewise-linear methodology relies on very strong exogeneity assumptions. Other than hours of work, all variables involved in the calculation of taxes are presumed to be exogenous determinants of labor supply behavior, both from a statistical and from an economic perspective. These variables include gross wages, the various components of non-labor income, and deductions. In light of the evidence supporting the view that wages and income are endogenous variables in labor supply analyses, particularly in the case of wages,[38] suspicions arise regarding the dependability of estimated substitution and income effects based on procedures that ignore such possibilities. Most of the exogeneity assumptions are also maintained in the differentiable-likelihood approach, but are easily relaxed when applying instrumental-variable procedures (given the availability of a sufficient number of other instrumental variables).

Fourth, some concerns about the reliability of estimates produced by the piecewise-linear approach ensue due to the static behavioral framework maintained in the formulation of empirical relations. Piecewise-linear studies invariably rely on the textbook one-period model of labor supply as a description of hours-of-work choices, and impose it to estimate parameters. Existing implementations of the differentiable-likelihood approach suffer from the same problem. Everyone acknowledges that individuals are not simply myopic optimizers; they transfer income across periods to achieve consumption plans that are infeasible without savings. A serious question arises concerning the relevance of such considerations in estimating substitution and income effects used to predict responses to tax policy.

6.4.4. Imposition of behavioral restrictions with convex constraints

The econometric model produced by this piece-wise linear formulation implicitly imposes parametric restrictions that constrain the signs of estimated substitution and income effects. As developed in MaCurdy et al. (1990), particular inequality restrictions must hold in the application of estimation procedures with piecewise-linear budget constraints for likelihood functions to be defined (i.e., to ensure that the components of these functions are non-negative). More specifically, in applications of such procedures the Slutsky condition must be locally satisfied at all interior kink points of budget sets that represent feasible options for any individual in the sample such that the compensated substitution effect must be positive. For the linear specification of the labor supply function considered in the preceding discussion, the specific inequality constraints imposed are

$$\alpha - \beta H_{ji} \geq 0, \quad \forall i, j, \tag{6.29}$$

where the quantities H_{ji} represent the hours-of-work values that correspond to interior kink

[37] See Blomquist (1996) for some Monte Carlo comparisons of ML and IV with measurement error.

[38] See Pencavel (1986) for a summary of this evidence.

points j on a sample member i's budget set. Because many values of H_{ji} exist in most analyses of piecewise-linear constraints, fulfillment of relations (6.29) essentially requires global satisfaction of the Slutsky condition by the labor supply function. Such a requirement, in essence, globally dictates that the uncompensated substitution effect of a wage change on hours of work must be positive for the labor supply specification considered in the preceding discussion, and the income effect for hours of work must be negative. The imposition of these restrictions, especially for men, is highly suspect given the available evidence from other studies. These restrictions carry over to more general labor supply functions.

6.5. *Maximum likelihood: accounting for fixed costs of participation and missing wages*

6.5.1. *Fixed costs*

As mentioned above, some applications of the piecewise-linear approach incorporate fixed costs to working – costs such as transportation that must be paid for any amount of work but which may vary across individuals. This significantly complicates the analysis because the optimized level of work under the budget constraint while working may not represent the optimal choice overall; one must explicitly consider the option of not working and thus avoiding the fixed costs. For any level of fixed costs, a minimum number of hours worked is implied creating an attainable range in the observable hours of work distribution; individuals will not work unless the gain is large enough to overcome the fixed costs. In essence, these complications arise because the budget constraint is not convex, invalidating simple maximization procedures.

If an individual must pay fixed monetary costs, F, to work, then non-labor income, Y, in the above budget constraints is replaced by

$$\begin{aligned} &Y - F, && \text{if } h > 0, \\ &Y, && \text{if } h = 0. \end{aligned} \tag{6.30}$$

F is partially unobservable and, thus, modeled as a stochastic element, varying across individuals. Hence, we see that the budget constraint discontinuously jumps down by F when the individual chooses to work.

To solve for the optimum when faced with this budget constraint, two regimes must explicitly be considered: working and not working. Estimation proceeds by finding the maximum utility under each regime and then comparing these to determine which option is chosen. In either regime, the utility function $U(C,h,v)$ – where we explicitly note the unobserved component, v – is maximized subject to Eq. (6.1) modified by Eq. (6.30).

In the no-work regime, the solution is simple. We know h is 0, so utility is given by $U(Y - \tau(Y - D), 0, v)$.

The solution in the work regime closely follows the solution presented in Section 6.3. Again utilizing the labor supply function, $f(\omega,y,v)$ yields the solution for h given in Eq (6.25), where the virtual income y now subtracts fixed costs F. However, to compute

maximum utility in this regime requires associating a utility level with each possible hours choice. Utility along any segment, i, is given by the indirect utility function, $V(\omega_j, y_j, \nu)$. At kinks, the direct utility function must be used, so the utility at kink j is given by $U(\omega_j H_j + y_j, H_j, \nu)$. Hence, utilizing exactly the same solution procedure derived in Section 6.3, we can define maximized utility when working, V^*:

$$
\begin{aligned}
&-\infty, && f_1 \leq 0,\\
&V(\omega_1, y_1, \nu), && 0 < f_1 < H_1,\\
&U(\omega_1 H_1 + y_1, H_1, \nu), && f_2 < H_1 \leq f_1,\\
&V^*(\omega, y, \nu) = V(\omega_2, y_2, \nu), && H_1 < f_2 < H_2,\\
&U(\omega_2 H_2 + y_2, H_2, \nu), && f_3 < H_2 \leq f_2,\\
&V(\omega_3, y_3, \nu), && H_2 < f_3 < H_3,\\
&U(\omega_3 \bar{H} + y_3, \bar{H}, \nu), && f_3 \geq \bar{H},
\end{aligned}
\tag{6.31}
$$

where

$$
f_j = f(\omega_j, y_j, \upsilon) \equiv \frac{V_\omega(\omega_j, y_j, \upsilon)}{V_y(\omega_j, y_j, \upsilon)}, \tag{6.32}
$$

with V_ω and V_y denoting the partial derviatives of V; relation (6.32) is, of course, Roy's identity defining the labor supply function, f, evaluated at wage and income levels ω_j and y_j. The use of $-\infty$ for $h = 0$ simply indicates that $h = 0$ is not included in this regime and, thus, selecting it indicates that the no-work regime is preferred. Given functional forms for V and U, finding V^* is straightforward.

Given maximized utility under each regime, the final step in the solution is to compare the two regimes. An individual chooses to work at the hours specified by the solution in Eq. (6.25) if

$$
V^*(\omega, y, \nu) \geq U(Y - \tau(Y \quad D), 0, \nu) \tag{6.33}
$$

and chooses not to work otherwise. For any level of ν, treating Eq. (6.33) as an equality implies a critical level of fixed costs, $F^*(\nu)$ above which the individual will choose not to work; F enters this relation through the virtual income variable y. Because desired hours of work increase with ν, this critical value will generally be increasing in ν – greater propensity to work implies that higher fixed costs are required to prefer the no-work option. If restrictions are placed on the support of F, such as $F > \underline{F}$, there will be values of ν low enough to rule out the work regime, thus implying a hole at the low end of the h distribution.

6.5.2. Missing wages

As a final step before deriving the likelihood function, note that in the no-work regime, gross wage, W, is not observed and, thus, the budget constraint cannot be derived. Hence, W must be endogenized. This can be accomplished by the simple function

$$W = W(Z) + \eta \tag{6.34}$$

where Z includes all observable variables determining W and η is the unobservable component. In a richer model, the equation for W could be derived as an equilibrium condition.

To derive the likelihood function, first consider the likelihood contribution of an individual who does not work. We assume this no-work decision can be observed, so there is no measurement error. In the no-work case, one of two situations applies: (i) fixed costs are sufficiently high with $F > F^* \equiv F^*(\nu, \eta)$ for any given ν and η, or (ii) if this fixed-cost threshold falls below the lowest admissible value for F (i.e., $F^* \leq F$), then desired hours are sufficiently low with $\nu < \nu^* \equiv \nu^*(\eta)$ for any η.[39] The probability of this event is

$$l_0 = \int_{-\infty}^{\infty} \int_{-\infty}^{\nu^*} \int_{F^*}^{\infty} g_{\nu\eta F}(\nu, \eta, F) \mathrm{d}F \mathrm{d}\nu \mathrm{d}\eta, \tag{6.35}$$

where $g_{\nu\eta F}$ is joint density of (ν,η,F).

For the work regime, the likelihood contribution looks very much like that derived in Eq. (6.28), as we continue to assume the linear hours of work function and the form of measurement error assumed there. The only changes are the addition of terms for δ and F (accounting for the fact that $F < F^*(\nu)$) and the removal of the term for the lower limit which is no longer part of that regime and is now perfectly observable. Using g_1 and g_2 to denote the distribution of (ε,ν,η,F) and $(\varepsilon + \nu,\nu,\eta,F)$ yields

$$l_1 = \sum_{j=1}^{3} \int_{\nu_j}^{\bar{\nu}_j} \int_0^{F^*} g_2[h^* - f_j, \nu, W - W(Z), F] \mathrm{d}F \mathrm{d}\nu + \sum_{j=1}^{2} \int_{\bar{\nu}_j}^{\underline{\nu}_{j+1}} \int_0^{F^*} g_1[h^* - H_j, \nu, W$$

$$- W(Z), F] \mathrm{d}F \mathrm{d}\nu + \int_{\bar{\nu}_3}^{\infty} \int_0^{F^*} g_1[h^* - \bar{H}, \nu, W - W(Z), F] \mathrm{d}F \mathrm{d}\nu, \tag{6.36}$$

where

$$\underline{\nu}_j \text{ solves the equation } f(\omega_j, y_j, \underline{\nu}_j) = H_{j-1},$$

$$\bar{\nu}_j \text{ solves the equation } f(\omega_j, y_j, \bar{\nu}_j) = H_j. \tag{6.37}$$

All variables are defined as in Section 6.4. Define $P_{ij} = 1$ if the individual works and 0 otherwise. Then the likelihood function for an individual is given by

[39] The critical value ν^* solves relation (6.33) treated as an equality with virtual income y evaluated at F.

$$l = (l_1)^{P_E}(l_0)^{1-P_E}. \tag{6.38}$$

Estimation proceeds by maximizing the sum of log likelihoods across individuals, as always.

This is quite complex in this case, requiring knowledge of both the direct utility U and the indirect utility V, and also requiring comparisons across regimes for all individuals and all parameter values.

6.6. Welfare participation and non-convex budget constraints

A common source of non-linearity in budget constraints involves participation in welfare programs. To illustrate this situation, consider the simplest case in which the only taxes faced by an individual result from benefit reduction on a single welfare program. Fig. 18 presents this scenario. As discussed in Section 2 of this survey, individuals face very high effective tax rates when they initially work due to large reductions in their benefits occurring when earnings increase. Once benefits reach 0, the tax rate drops to a lower level, creating a non-convex kink in the budget constraint. This non-convexity invalidates the simple procedures of Section 6.4 implemented to divide sample spaces into locations on budget sets.

6.6.1. Simplest welfare case with no stigma

In this simple case, an individual maximizes $U(C,h,v)$ subject to the budget constraint

$$C = Wh + Y + B(I(h)), \tag{6.39}$$

where benefits are given by the simplest benefit schedule:

$$B(I(h)) = \begin{cases} G - bWh, & \text{if } G - bWh > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{6.40}$$

G gives the guarantee amount which is reduced at the benefit reduction rate b as the earnings, Wh, increase. This implies a kink point at $H_1 = G/bW$ where benefits reach 0 and, thus, the marginal wage rises to W. So, the individual faces two segments: segment 1 has $h < H_1$ with net wage $\omega_1 = (1 - b)W$ and virtual income $y_1 = Y + G$; and segment 2 has $h > H_1$ with net wage $\omega_2 = W$ and virtual income $y_2 = Y$.[40]

Because the budget constraint is non-convex, the solution cannot be characterized simply by finding a tangency with the budget constraint as it was in Section 6.3. Multiple tangencies are possible and these must be directly compared to determine the optimum. Hence, the regime shift approach of Section 6.5 is needed.

Consider first the regime in which positive benefits are received; that is, $h < H_1$. Maximization, given the effective wage and income, on this linear segment follows the

[40] We continue to use N to denote unearned non-taxable income for ease of notation. In addition, we ignore any upper bound on hours worked for simplicity.

method of Section 6.3. We can characterize the optimal choice according to the function $f(\omega_1, y_1, \nu)$. Denote the value of ν which implies $f(\omega_1, y_1, \nu) = 0$ as ν_0. Then the optimal hours choice along that segment is given by

$$h = f(\omega_1, y_1, \nu), \qquad \nu > \nu_0, \qquad h = 0, \qquad \nu \leq \nu_0. \tag{6.41}$$

The optimized value on this segment (including the zero work option), accounting for the fact that $h > H_1$ is not allowed, is given by

$$V_1^*(\omega_1, y_1, \nu) = \begin{cases} V(\omega_1, y_1, \nu), & 0 < f_1 \leq H_1 \\ U(y_1, 0, \nu), & f_1 \leq 0 \\ -\infty, & f_1 > H_1. \end{cases} \tag{6.42}$$

where Eq. (6.32) defines f_1.

Next, consider the regime without benefits, that is with $h \geq H_1$. Again the optimal choice, given the wage and income, on this segment is given by the labor supply function $f(\omega_2, y_2, \nu)$. The optimized value, accounting for the fact that $h < H_1$ is not admissible, is given by[41]

$$V_2^*(\omega_2, y_2, \nu) = \begin{cases} V(\omega_2, y_2, \nu), & f_2 \geq H_1, \\ -\infty, & f_2 < H_1. \end{cases} \tag{6.43}$$

Hence, the individual selects regime 1, with welfare receipt, if $V_1^* > V_2^*$, and regime 2 otherwise. Since work propensity increases with ν, this can be characterized by a cutoff value, ν^*, defined by

$$V_1^*(\omega_1, y_1, \nu^*) = V_2^*(\omega_2, y_2, \nu^*). \tag{6.44}$$

For values of ν above ν^*, regime 2 is chosen; and for values below ν^*, regime 1 is realized.

We can define three sets, Ω_0, Ω_1, and Ω_2, such that for $\nu \in \Omega_0$ the individual chooses not to work, for $\nu \in \Omega_1$ the individual locates on segment 1 with positive hours of work, and for $\nu \in \Omega_2$ the individual locates on segment 2. We must consider two cases to define these sets exactly. First, suppose $\nu^* > \nu_0$. Then we have

$$\Omega_0 = \{\nu \mid \nu \leq \nu_0\},$$

$$\Omega_1 = \{\nu \mid \nu_0 < \nu \leq \nu^*\},$$

$$\Omega_2 = \{\nu \mid \nu > \nu^*\}. \tag{6.45}$$

Alternatively, if $\nu^* \leq \nu_0$, then the switch to regime 2 occurs before positive hours are worked in regime 1, that is

[41] In the following formulation, we implicitly assume that the event $f_2 \geq H$ occurs with zero probability

$$\Omega_0 = \{\nu \mid \nu \leq \nu^*\},$$

$$\Omega_1 = \varnothing,$$

$$\Omega_2 = \{\nu \mid \nu > \nu^*\}. \tag{6.46}$$

Hence, for certain individuals and parameter values, no value of ν exists such that they will locate on segment 1 with positive hours of work.

To characterize the likelihood function we again need a functional form for the gross wage of the form $W = W(Z) + \eta$. We ignore measurement error here for simplicity, and because there is no problem with individuals failing to locate at the kink in this non-convex case. Define $P_B = 1$ if the individual receives benefits, and $P_E = 1$ if the individual works, both 0 otherwise. The likelihood function is given as follows, incorporating $g_{\eta\nu}(\eta,\nu)$ and the general inverse function $\nu = \nu(h)$:

$$P_B = 1, \qquad P_E = 1, \qquad l_{11} = \frac{\partial \nu}{\partial h} g_{\nu\eta}(\nu(h), W - W(Z)) I(\nu \in \Omega_1),$$

$$P_B = 0, \qquad P_E = 1, \qquad l_{01} = \frac{\partial \nu}{\partial h} g_{\nu\eta}(\nu(h), W - W(Z)) I(\nu \in \Omega_2),$$

$$P_B = 1, \qquad P_E = 0, \qquad l_{10} = \int_{\Omega_0}^{\infty} g_{\nu\eta}(\nu, \eta) \mathrm{d}\nu \mathrm{d}\eta, \tag{6.47}$$

where $I(\cdot)$ represents an indicator function equal to 1 if the condition in the parentheses is true. Because the value of ν implied by the hours choice may be inconsistent with the value implied by the regime choice, it is possible to have "holes" in the hours distribution around the kink point. For example, an individual on segment 1 must have $\nu \leq \nu^*$. If his hours choice is too close to the kink, this may imply a value of $\nu > \nu^*$ and thus an observation with zero likelihood.

The overall likelihood function is given by

$$l = (l_{11})^{(P_B)(P_E)} (l_{01})^{(1-P_B)(P_E)} (l_{10})^{(P_B)(1-P_E)}. \tag{6.48}$$

Estimation proceeds by maximizing this sum of the log likelihoods across individuals.

6.6.2. *Welfare stigma*

The above analysis assumes that all individuals eligible for welfare are on welfare. Individuals working less than h_0 but failing to receive welfare are operating below the implied budget constraint, a possibility not permitted in the analysis. Yet, many individuals are in exactly this situation. This is generally explained by assuming the existence of some utility loss or stigma associated with welfare.

To capture welfare stigma the utility function is modified to take the form

$$U = U(C, h, \nu) - P_B \xi, \tag{6.49}$$

where ξ is the level of welfare stigma which is greater than 0 and varies across individuals.[42] With this modification we again consider the welfare and non-welfare regimes. Since the welfare stigma term does not affect the marginal decisions, given that the individual is on welfare, the discussion of hours of work presented above for regime 1 is still valid. The optimal utility is now given by

$$V^*(\omega_1, y_1, \nu) = \begin{cases} V_1(\omega_1, y_1, \nu) - \xi, & 0 < f_1 \leq H_1, \\ U(y_1, 0, \nu) - \xi, & f_1 \leq 0, \\ -\infty, & f_1 > H_1. \end{cases} \tag{6.50}$$

The analysis for regime 2 is altered in this case, because an individual can be observed on welfare for any value of h – that is, given welfare stigma, it is possible to observe an individual with $h < H_1$, but $P_B = 0$. So regime 2 is now defined solely by $P_B = 0$. Optimal hours of work, given ω_2 and y_2, are given by $f(\omega_2, y_2, \nu)$. Defining the value of ν for which $f(\omega_2, y_2, \nu) = 0$ as ν^+, hours of work under this regime are now given by

$$h = f(\omega_2, y_2, \nu), \qquad \nu > \nu^+,$$

$$h = 0, \qquad \nu \leq \nu^+. \tag{6.51}$$

Optimized utility is now

$$V_2^*(\omega_2, y_2, \nu) = \begin{cases} V(\omega_2, y_2, \nu), & f_2 > 0 \\ U(y_2, 0, \nu), & f_2 \leq 0. \end{cases} \tag{6.52}$$

Choice of regime still proceeds by comparing V_1^* and V_2^*, as done in Eq. (6.44). For any ν in the sets Ω_0 or Ω_1 defined by Eq. (6.45) or (6.46), there is now some critical level of $\xi^* \equiv \xi^*(\nu)$, which depends on ν, such that regime 2 is chosen when $\xi > \xi^*$; regime 1 is chosen otherwise.

Given this characterization, we can derive the likelihood function for each combination of P_B and P_E, using the joint densities $g_{\nu\xi\eta}(\nu, \xi, \eta)$ and $g_{\nu\eta}(\nu, \eta)$:

$$P_B = 1, \qquad P_E = 1, \qquad l_{11} = \frac{\partial \nu}{\partial h} \int_0^{\xi^*} g_{\nu\xi\eta}(\nu(h), \xi, W - W(z)) I(\nu \in \Omega_1) \mathrm{d}\xi,$$

$$P_B = 0, \qquad P_E = 1, \qquad l_{01} = \frac{\partial \nu}{\partial h} g_{\nu\eta}(\nu(h), \xi, W - W(z)) I(\nu \in \Omega_1)$$

$$+ \frac{\partial \nu}{\partial h} \int_{\xi^*}^{\infty} g_{\nu\xi\eta}(\nu(h), \xi, W - W(z)) I(\nu \in \Omega_1) \mathrm{d}\xi,$$

[42] This additive form is used for simplicity. More general forms can be used, but change none of the substantive points presented here.

$$P_B = 1, \qquad P_E = 0, \qquad l_{10} = \int_{-\infty}^{\infty} \int_{\Omega_0} \int_{0}^{\xi^*} g_{\nu\xi\eta}(\nu, \xi, \eta) \mathrm{d}\xi \mathrm{d}\nu \mathrm{d}\eta,$$

$$P_B = 0, \qquad P_E = 0, \qquad l_{00} = \int_{-\infty}^{\infty} \int_{-\infty}^{\nu^+} \int_{0}^{\xi^*} g_{\nu\xi\eta}(\nu, \xi, \eta) \mathrm{d}\xi \mathrm{d}\nu \mathrm{d}\eta. \tag{6.53}$$

Estimation proceeds as in the non-stigma case by selecting the appropriate likelihood branch for each individual and then maximizing the sum of the log likelihoods.

As with the fixed cost case, the likelihood function is complex even in this extremely simplified welfare case. For each possible set of parameter values, the maximum must be computed for each regime and then compared to compute ξ^*. Adding the tax codes, with their implied kinks, increases computational complexity. As a result, the literature has adopted a simplifying methodology which we present in Section 6.8.

6.6.3. Multiple program participation

In principle, the extension to the case of multiple program participation is straightforward. For simplicity, we consider a case in which the individual can choose between participating in no welfare programs, participating in welfare program 1, participating only in program 2, or participating in both welfare programs 1 and 2. We extend the utility function as follows:

$$U = U(c, h, \nu) - P_1\xi - P_2\chi \tag{6.54}$$

where $P_1 = 1$ if the individual participates in program 1, and $P_2 = 1$ if the individual participates in program 2.[43] Benefits from program 1, $B_j(I(h))$, are given:

$$B_j(I(h)) = \begin{cases} G_j - b_j Wh, & \text{if } G_j - b_j Wh > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{6.55}$$

Benefits from both together are given as

$$B_1(I(h)) + B_2(I(h)) = \begin{cases} G_1 + G_2 - b_1 Wh - b_2 Wh = G - bWh, & \text{if } G - bWh > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{6.56}$$

where $G = G_1 + G_2$ and $b = b_1 + b_2$. In general, the benefit functions for programs 1 and 2 will have different breakeven points, implying the values of hours defining kinks (H_1 in Fig. 18) will not be the same.

This formulation expands the model considered in Section 6.4.3. To adapt this earlier model, one must designate three distinct regimes in place of regime 1 specified above: regime 1a indicating an individual participates only in program 1, regime 1b signifying this person collects benefits only from welfare program 2, and regime 1c designating

[43] The use of two additive errors is a simplifying assumption which ensures that the stigma from both programs is higher than stigma from program 1 alone.

participation in both programs. Optimal hours and utility for participation in a regime are given by (6.41), (6.42), (6.50), (6.51), and (6.52), with net wages and virtual income in these formulations specified as $\omega_j = W(1 - b_j)$ and $y_j = Y + G_j$, with $j =$ 1a, 1b, or 1c. In particular, relations analogous to (6.41) and (6.42) define the labor supply and utility functions for each of the new regimes for the "on-welfare" segments associated with relevant combination of welfare programs. Relations (6.51) and (6.52) still define the labor supply and utility functions for the non-welfare regime. The set of relations define thresholds for ν demarcating the regions of unobserved tastes determining when a person works (ν_0 in (6.41) and ν^+ in (6.51)). Maximization again requires selection of a regime. Relations analogous to (6.50) and (6.52) characterize utilities corresponding to the various regimes. Conditional on values ν, these relations in turn imply thresholds for the stigma errors ξ, χ, and $\xi + \chi$ that determine individuals' welfare participation. The likelihood function for this model takes a form similar to Eq. (6.53), with more branches appearing in the function reflecting the additional regimes analyzed in this formulation.

Again, note the complexity of these, extremely simplified welfare cases, even these involve significantly financial burden. For each possible set of parameter values, one must compute the maximum for each regime, account for the benefit structure, and then compare these to compute the error ranges for the likelihood function. When the individual is unemployed, one must perform these calculations for all possible wage values and all values of ν consistent with the no-work decision. Adding the tax code, with its implied kinks, increases computational difficulties. Introducing additional sources of unobserved heterogeneity enlarges the number of dimensions over which one must calculate integrals, requiring sophisticated numerical procedures and considerable computer resources. As a result, the literature has adopted simplifying methodologies, a topic to which we now turn.

6.7. An approach for computational simplification and discrete hours choices

To make estimation problems manageable, a popular method is to presume that consumers face only a limited set of hours choices. For example, a worker may choose only full-time work, part-time work, or no work, with each of these options implying a prescribed number of hours. Formally, this is done by assuming that unobservable tastes components, ν, possess a discrete distribution, usually characterized as a multinomial distribution conditional on covariates. Combined with a 0/1 welfare decision, this finite set of hours choices yields a relatively small set of discrete states, say S states, over which the utility function must be maximized.

Given a specific form for the preference function, utility can be readily evaluated at each of the hours choices and the maximum can be determined. Given an assumed joint distribution for unobservable tastes components, ν, for the error component determining wages, η, and for welfare stigma, ξ, one can compute a probability that a family selects alternative "j". This in turn defines a sample log likelihood of the form

$$l = \sum_{j=1}^{S} d_j \ln P(j \mid X, \Theta), \tag{6.57}$$

where d_j is an indicator for whether individual i chooses alternative j, X is a vector of observable characteristics, and $P(j \mid X, \Theta)$ is the probability of choosing alternative j with Θ the set of unknown parameters. Such formulations are substantially less complicated than the specifications considered above because one avoids the intricate process of calculating thresholds and dealing with combined continuous-discrete endogenous variables; only discrete choices are allowed for here.

This formulation requires each individual to be placed into a limited set of preassigned work states, even though observed hours worked take many more values, making hours look as if they were continuously distributed. To overcome this issue, analyses applying this approach necessarily introduce measurement error in hours of work to admit hours to deviate from the discrete values assumed for the choice set. Hence, conditional on ν, each alternative "j" contributes some positive probability $P(j \mid X, \Theta, \nu)$ which now depends on the value of the unobservable measurement error variables.

We illustrate this approach by considering the linear measurement error model given by Eq. (6.19) where the reporting error $\varepsilon \sim g_\varepsilon$, with ε and υ independent. Further, as typically assumed, we specify that hours are not subject to measurement error in no-work states. The likelihood function for hours now takes the form

$$l = \left(\sum_{j \in S_0} d_j \ln P(j \mid X, \Theta) \right)^{1 - P_E} \left(\sum_{j \in S_1} d_j \ln(g_\varepsilon(h - h_j) P(j \mid X, \Theta)) \right)^{P_E}, \tag{6.58}$$

where P_E denotes a 0/1 variable with 1 indicating that the individual works, S_0 designates the set of all states associated with the individual not working, the set S_1 includes all states in which the individual works, and h_j denotes the admissible values of true hours. Earnings depend on the values of h_j and wages. In Eq. (6.58), observed hours are continuously distributed among workers.

6.8. *Survey of empirical findings for non-linear budget constraints models*

Having developed a theoretical framework for analyzing the effects of taxes on labor supply, we proceed in this section to briefly survey the body of empirical literature that seeks to estimate labor supply elasticities in the presence of welfare and taxes. The selection of studies considered here illustrate the empirical methodologies developed in the preceding subsections.[44] The survey begins with the empirical work involving maxi-

[44] See also Burtless and Hausman (1978), Heckman (1979c), Hausman (1980, 1985a,b), Cogan (1981), Nakamura and Nakamura (1981), Ashenfelter (1983), Moffitt (1983, 1986, 1992a,b), Fraker et al. (1985), Robins (1985), Blundell et al. (1986, 1988), Blau and Robins (1988), Fraker and Moffitt (1988), Zabalza (1988), Kell and Wright (1989), Moffitt and Wolfe (1992), Ribar (1992), Bingley et al. (1995), Blomquist (1996) and Bingley and Walker (1997).

mum likelihood estimation with convex budget sets. This case, which we discussed in Sections 6.3 and 6.4, has received the most attention in the received empirical literature. We then consider papers that report at least one model estimated by instrumental variables or involving non-convex budget sets, the cases we considered in Sections 6.2 and 6.5. The survey concludes with an application of the multiple welfare program participation model which we presented in Section 6.6. In what follows we restrict our attention to post-1980 analyses of the United States and Western Europe. Tables 1 and 2 summarize the results for men and women, respectively.

6.8.1. Maximum likelihood estimation with convex budget sets

Blomquist (1983) estimates labor supply functions for prime-age Swedish males, using a piecewise-linear analysis with a convex budget set to account for the highly progressive Swedish income tax. His approach follows closely that of Hausman (1981). The 1973 cross section of 688 males, aged 25–55, used in his study is derived from a survey conducted by the Swedish Institute for Social Research. Estimation is based on the following linear model:

$$h_i^* = \alpha\omega_i + \beta y_i + \gamma Z + \varepsilon_i, \tag{6.59}$$

where α, β, and γ are preference parameters, h_i^* is hours worked in 1973, Z is a vector of individual characteristics, ω_i is the net wage rate on the ith extended budget segment, y_i is the virtual income for this segment, and ε_i is a disturbance. Non-labor income is defined as the spouse's after-tax income plus the family's capital income after tax and family allowances, where after-tax capital income is computed as it would have been if the person had worked no hours.

Blomquist assumes that α and γ are constant across individuals, whereas each person's β is assumed to be a draw from $f(\beta)$, the normal density function with upper truncation at zero. Since the individual β_i are not identified, Blomquist estimates μ_β and σ_β^2, the parameters of f, in addition to α and γ. Estimation by maximum likelihood yields an income elasticity of -0.03, a compensated wage elasticity of 0.11, and an uncompensated wage elasticity of 0.08. The author also reports results from estimation with the restriction $\sigma_\beta^2 = 0$ imposed. Although the resulting estimates are similar to those of the unconstrained model, a likelihood ratio test rejects the restriction at conventional levels.

MaCurdy et al. (1990) analyze the labor supply of prime-age married males using the 1975 cross-section from the Michigan Panel Study of Income Dynamics (PSID). They show that maximum likelihood estimation of a consumer-choice problem with non-linear budget sets implicitly relies on the satisfaction of inequality constraints that translate into behaviorally meaningful restrictions. These constraints arise from the requirement to create a well-defined statistical model, and not as a consequence of economic theory (see Section 6.4.4). The authors then present empirical results suggesting that these implicit constraints play a major role in explaining the disparate results found in the literature on men's labor supply. The empirical work is based both on the piece-wise linear approach

Table 1
Non-linear budget constraint models: summary of some empirical results for men

Study	Data source and sample selection	Variables: H, hours: W, wage: Y, income	Functional form of labor supply and budget set structure	Estimation method and stochastic specification[a]	Uncompensated wage elasticity	Income elasticity
Blomquist (1983)	Swedish Level of Living Survey 1974: sample size 688, all employed, married, aged 25–55	H, annual hours for 1973 (weeks worked × average hours per week) W, directly observed Y, spouse's net income ± family allowances + net capital income	Linear labor supply, convex (piecewise linear)	ML ML random preferences (on income coefficient)	0.08 0.08	−0.03 −0.04
Blomquist and Hansson-Brusewitz (1990)	Swedish Level of Living Survey 1981: sample size 602, all employed, married, aged 25–55	H, annual hours W, directly observed Y, spouse's net income + family allowances + net capital income	Linear and quadratic labor supply Convex and non-convex (piecewise linear)	*Linear labor supply* ML-convex ML-non-convex ML-convex random preferences *Quadratic labor supply* ML-convex	0.08 0.08 0.13 0.12	0.002 −0.008 −0.01 −0.008
Bourguignon and Magnac (1990)	French Labour Force Survey 1985: sample size 1992, all employed, married, aged 18–60	H, normal weekly hours W, hourly net wage (monthly earnings / hours) Y, family allowances	Linear labor supply Convex (piecewise linear)	ML-convex, random preferences	0.1	−0.07
Blundell and Walker (1986)	British Family Expenditure Survey 1980: sample size 1378, all employed, married, aged 18–59	H, usual weekly hours W, weekly earnings/ hours Y, consumption based two-stage budgeting [b]	Gorman polar form/ translog Convex (piece-wise linear)	ML-convex, random preferences	0.024	−0.287
Flood and MaCurdy (1992)	Swedish Household Market and Non-market Survey (HUS)	H, annual hours W, hourly wage (annual earnings/	Linear and semi-logarithmic Convex (piecewise	*Linear labor supply* ML-piecewise linear, random preferences	0.16	−0.1

	1984: sample size 492, all employed, married, aged 25–65	annual hours) Y, asset income, UI, housing allowances etc.	linear and differentiable)	(on income coefficient) additive measurement error		
				ML-differentiable, random preferences (on income coefficient) measurement error:		
				Additive	0.14	−0.09
				Multiplicative	0.04	−0.07
				None	0.07	−0.08
				Semi-log labor supply		
				ML-differentiable, random preferences, measurement error:		
				Multiplicative	0.21	−0.09
				None	0.25	−0.1
				IV across 7 different specifications	{−0.25, 0.21}	{−0.11, 0.04}
Hausman (1981)	US Panel Study of Income Dynamics 1975: sample size 1085, all employed, married, aged 25–55	H, annual hours W, directly reported hourly wage rates Y, other income assuming 8% return to financial assets	Linear labor supply	ML, random preferences (on income coefficient)	{0.00, 0.03}	{−0.95, −1.03}
Kaiser et al. (1992)	German SocioEconomic Panel 1983: sample size 2382 employed, 939 non-employed, married, non-retired	H, annual hours W, hourly wage (annual income/annual hours) Y, income from rents, capital income and transfer payments	Convex and non-convex (piecewice-linear)	ML-convex Linear labor supply convex (piecewise linear)	−0.004	−0.28

Table 1 (*continued*)

Study	Data source and sample selection	Variables: H, hours; W, wage; Y, income	Functional form of labor supply and budget set structure	Estimation method and stochastic specification[a]	Uncompensated wage elasticity	Income elasticity
MaCurdy et al. (1990)	US Panel Study of Income Dynamics 1975: sample size 1017, all employed, married, aged 25–55	H, annual hours W, average hourly wage (earnings/annual hours) Y, rent, interest, dividends, etc.	Linear labor supply Convex and non-convex (piecewise-linear and diffrentiable budget constraints)	ML-convex, random preferences[c] (on income coefficient)	0	−0.01
Triest (1990)	US Panel Study of Income Dynamics 1983: sample size 978, all employed, married, aged 25–55	H, yearly hours in all jobs held in 1983 W, average hourly earnings (earnings/annual hours) Y, rents, dividends, interest income, trust funds, etc.	Linear labor supply Convex (piecewise linear)	ML-convex, random preferences[c] (on income coefficient)	0.05	0[d]
van Soest et al. (1990)	Dutch Strategic Labor Market Research Survey 1985: sample size 801 employed, 49 non-employed	H, average weekly hours W, net hourly wage (earnings/hours) Y, other incomes	Linear Labor Supply Convex (piecewise linear)	ML-convex	0.12	−0.01

[a] This column indicates whether preferences are treated as random in specifications, in addition to "optimization" or measurement errors that are always incorporated in specifications. Unless stated otherwise, random preferences without further indications means that only intercept coeffcients are stochastic. For further information, see Eq. (6.59) and related discussion.

[b] See Eq. (4.33) and related discussion.

[c] For other results see article.

[d] Estimated coefficient constrained at zero.

Table 2
Non-linear budget constraint models: summary of some empirical results for married women

Study	Data source and sample selection	Variables: H, hours; W, wage[a]; Y, income	Functional form of labor supply and budget set structure	Estimation method and stochastic specification[b]	Uncompensated wage elasticity	Income elasticity
Arellano and Meghir (1992)	British Family Expenditure Survey (FES) 1983 and British Labor Force Survey (LFS) 1983: sample size 11,535 employed, 13,200 non-employed, aged 20–59	H, weekly hours W, hourly earnings[c] (earnings/hours) Y, consumption based other income measure	Semi-log labor supply Convex (piecewise linear)	Instrumental variables/selection	{0.29, 0.71}	{−0.13, −0.40}
Arrufat and Zabalza (1986)	Bristish General Household Survey 1974: sample size 2002 employed, 1493 non-employed, aged < 60	H, weekly hours W, gross hourly earnings, SS Y, net weekly unearned family income + husband's earnings	CES utility based labor supply Convex (piecewise linear)	ML-convex, random preference (log normal on CES leisure coefficient)[d]	2.03	−0.2
Blomquist and Hansson-Brusewitz (1990)	Swedish Level of Living Survey 1981: sample size 795 full sample, 640 employed, aged 25–55	H, annual hours W, directly observed, SS[e] Y, spouse's net income + family allowances + net capital income	Linear and quadratic labor supply Convex and non-convex (piecewise linear)	*Linear labor supply* ML-non-convex ML-non-convex, random preferences (on income coefficient) *Quadratic Labor Supply* ML-convex	0.79 0.77 0.58	−0.24 −0.06 −0.05
Blundell et al. (1988)	British Family Expenditure Survey 1980: sample size 1378 employed, aged 18–59	H, usual weekly hours W, hourly earnings[f] (earnings /hours) Y, consumption based two-stage budgeting measure	Generalized linear expenditure system Convex (piecewise linear)	Truncated ML, random preferences	0.09	−0.26

Table 2 (*continued*)

Study	Data source and sample selection	Variables: H, hours; W, wage[a]; Y, income	Functional form of labor supply and budget set structure	Estimation method and stochastic specification[b]	Uncompensated wage elasticity	Income elasticity
Bourguignon and Magnac (1990)	French Labor Force Survey 1985: sample size 1175 employed, 817 non-employed, aged 18–60	H, normal weekly hours W, hourly net wage, SS, (earnings /hours) Y, spouse's net income + family allowances	Linear labor supply Convex (piecewise linear)	ML-convex, random preferences ML with fixed costs, random prefrences	1 0.05	−0.3 −0.2
Colombino and Del Boca (1990)	Turin Survey of Couples 1979: sample size 338 employed, 494 non-employed	H, yearly hours (weeks worked × average weekly hours) W, hourly wage, SS, (annual earnings/ annual hours) Y, total net non-labor	Linear labor supply Convex (piecewise linear)	ML-convex	{1.18, 0.66}	0.52
Hausman (1981)	US Panel Study of Income Dynamics 1975: sample size 575 participants, 510 non-participants	H, annual hours of work W, directly reported hourly wage rates, SS[g] Y, transfer and asset income with 8% return to financial assets	Linear labor supply Convex (piecewise-linear) and non-convex (fixed costs)	ML-convex, random preferences ML-fixed costs random preferences (on income coefficient)	0.995 0.906	−0.121 −0.132
Kaiser et al. (1992)	German SocioEconomic Panel 1983: sample size 1076 employed, 2284 non-employed, non-retired	H, yearly hours W, hourly wage, SS (annual earnings/ annual hours) Y, income from rents, capital income and transfer payments	Linear labor supply Convex (piecewise linear)	ML-convex	1.04	−0.18

Kuismanen (1997)	Finnish Labor Force Survey 1989 sample size: 1541 employed 485 non-employed aged 25–60	H, yearly hours W, hourly wage, SS. (annual earnings/ annual hours) Y, Income from rents, dividends, capital income, etc.	Semi-log labor supply Convex (piecewise linear)	*Survey data* ML-convex fixed preferences *Tax register data* ML-convex, random preferences	−0.01 0.01	0.27 0.11
Triest (1990)	US Panel Study of Income Dynamics 1983: sample size 715 employed, 263 non-employed, aged 25–55	H, yearly hours in all jobs W, average hourly earnings, SS,[h] (earnings/hours) Y, rents, dividends, interests, etc.	Linear labor supply Convex and piecewise linear	*Full sample* ML-convex, random preferences (on income coefficient) *Workers only* ML-convex, random preferences (on income coefficient)	0.97 0.28	−0.33 −0.17
van Soest et al. (1990)	Dutch Strategic Labor Market Research Survey 1985: sample size 331 participants, 470 non-participants	H, average weekly working hours W, net hourly wage, SS (earnings/hours) Y, other incomes	Linear labor supply Convex (piecewise linear)	M-convex	0.79	−0.23

[a] SS signifies that wages are predicted via linear selectivity adjusted regression, in this particular study selection bias was argued not to be important. In each of the studies for married women reported here, we leave readers to refer to the original source for details of identification strategy and included regressors.

[b] This column indicates whether preferences are treated as random in specifications, in addition to "optimization" or measurement errors that are always incorporated in specifications. Unless stated otherwise, random preferences without further indications means that only intercept coefficients are stochastic. For further information, see Eq. (6.59) and related discussion.

[c] Hourly earnings and consumption based other income are introduced to the LFS via an instrumental regression on the FES.

[d] See Eq. (6.62).

[e] In the results reported here, unobserved wages are predicted using selection adjusted wage equations except for the fixed preference linear labor supply case in which unobservable wages are integrated out of likelihood.

[f] Only truncated sample of workers used in estimation.

[g] Predicted wage used for those out of employment, selection bias is argued not to be important for this sample.

[h] The results we report for the "workers only" use observed wages; the author reports similar results using imputed wages.

and the differentiable constraint case, and the authors only consider convexified budget sets.

MaCurdy, Green, and Paarsch consider three specifications of the labor supply function, along with both an additive and a multiplicative structure for the measurement error term. The first is a linear labor supply function with substitution and income effects constant across individuals. The second assumes the substitution coefficient to vary across individuals, and the third allows the income effects to vary. The third approach is also taken by Hausman (1981) and Blomquist (1983). In the piece-wise linear equations, the authors use only the additive structure for measurement error rather than the multiplicative, since the latter implies the unattractive feature that earnings are observed without error. The evidence from all models suggests a strong influence of the implicit inequality restrictions invoked by the maximum likelihood procedure. This offers an explanation for the divergent results of previous research relying on various empirical methodologies.

Arrufat and Zabalza (1986) use British cross-sectional data on married women from the 1974 General Household Survey to estimate a model of female labor supply that reflects the joint decision on labor force participation and hours, the non-linear budget constraint created by income taxation, the effect of heterogeneous preferences, and the existence of optimization errors. These optimization errors cause agents' actual position on the budget constraint to differ from their preferred position. The structural model is based on the following CES family utility index defined over net family income, x, and wife's leisure, l:

$$u = [x^{-\rho} + \alpha l^{-\rho}]^{-(1/\rho)} \tag{6.60}$$

with error structure

$$\alpha = \exp[\beta Z - \xi], \qquad x/l = (x/l)^{*}\exp(\varepsilon), \tag{6.61}$$

where $(x/l)^{*}$ is the utility-maximizing income-leisure ratio, Z is a vector of personal characteristics, β is a vector of parameters, and (ξ,ε) is distributed bivariate normal $(0,0,\sigma_{\xi}^2,\sigma_{\varepsilon}^2,0)$. The budget constraint looks like Fig. 15 in (x,l) space. The maximum likelihood estimator yields an estimated elasticity of substitution of 1.21. The elasticities with respect to own wages, husband's wages, and unearned family income are 2.03, -1.27, and -0.20. This own wage elasticity of approximately two is larger than those estimated in previous studies using British data.

Blundell et al. (1988) estimate a generalized version of the Stone–Geary labor supply model using a sample of almost 1400 married women from the British Family Expenditure Survey for 1980. A truncated likelihood approach was used that considered hours of work conditional on participation. The preference specification was chosen according to standard likelihood diagnostics.[45] Although uncompensated wage elasticities were small, the compensated elasticities were found to be quite large and positive across a wide range of demographic groups. This model was then used to simulate a number of reforms to the British tax system in Blundell et al. (1988).

[45] See Blundell and Meghir (1986).

Friedberg (1995) analyzes data from the United States March Current Population Survey. She uses a convex budget set with a piece-wise linear constraint for studying progressive taxes and the social security earnings test, and assumes a linear functional form for the labor supply equation. The Heckman sample selection technique is used to predict non-participant wages in the labor supply equation. Maximum likelihood estimates of the model yield a compensated wage elasticity of 1.12, an uncompensated wage elasticity of 0.36, and an income elasticity of -0.76.

Van Soest et al. (1990) analyze a cross-section of Dutch households from a 1985 labor mobility survey by the Organization of Strategic Labor Market Research. They consider a piece-wise linear framework with a convex budget set and normally distributed random preferences and optimization errors. As a second specification, they estimate a simple reduced form model of the demand side of the labor market, in which employers offer wage-hours packages and individuals choose among a limited number of these offers. The authors impose the distributional assumptions stated following Eq. (6.61), and estimate the models using maximum likelihood. In their second specification, the error term ν is replaced by a job offer mechanism, which treats the number of hours worked as a discrete rather than a continuous random variable. Their results imply wage-rate elasticities of 0.65 and 0.79 for women and 0.12 and 0.10 for men. These and the estimated income elasticities are in harmony with previous work using Dutch data.

6.8.2. Non-convex budget sets: maximum likelihood and instrumental variable estimation

Hausman (1981) estimates the effect of taxation and transfers on the labor supply of a subsample of prime-age husbands, wives, and female family heads who have children under the age of eighteen from the 1975 PSID, treating the husband as the primary earner and the wife as the secondary earner. For husbands and wives he considers two cases: the non-convex piece-wise linear case representing a tax and transfer schedule based on actual law, and a convexified tax schedule where the effects of FICA, the earned income credit, and the standard deduction are approximated by a consistently progressive convex budget set. For female household heads he considers only the non-convex case because of the large initial non-convexity introduced by AFDC.

Hausman assumes a linear functional form for the labor supply equation. Although the wage coefficient in the hours equation is assumed to be constant across individuals, the coefficient of virtual income is assumed to vary. Blomquist (1983) also uses this approach, as discussed at the beginning of this survey. Hausman assumes that the coefficient of virtual income is the mean of the truncated normal distribution. Since it is assumed that this coefficient is non-positive, the relevant part of the distribution is to the left of zero. Hausman considers the possibility of selection bias, since market wages are unobserved for non-workers, but finds that it is not a problem in his sample. Estimation is by maximum likelihood.

For husbands, he finds that the uncompensated wage coefficient is essentially zero which accords with previous empirical findings. However, his finding of a significant income effect is at odds with prior work. Since the wage and income variables from the

convex and non-convex budget sets are similar, Hausman concludes that for estimation purposes it is probably reasonable to smooth the non-convexities created by the earned income credit, social security taxes, and the standard deduction. For wives, he finds substantial uncompensated wage and income elasticities. In addition to the convex and non-convex cases, a specification that explicitly accounts for the fixed costs of working is included for wives. The resulting wage elasticities are midway between those of husbands and those of wives.

Triest (1990) considers the sensitivity of Hausman's results to changes in the model specification. To this end, he estimates several variants of Hausman's model using a 1983 subsample of the PSID. Both the labor supply equation and the measurement error equation are linear, with the distributional assumptions stated following Eq. (6.61). A specification representing preference heterogeneity as a random income coefficient, rather than an additive disturbance, is also estimated following Hausman. Triest considers maximum likelihood estimation under the assumptions of preference heterogeneity only, measurement error only, and both heterogeneity and measurement error, in addition to instrumental variables estimation assuming only heterogeneity. In the heterogeneity-only model for women, GMM was used to estimate an IV version of the Tobit model. Triest follows Hausman by treating the convex hull of the budget set as the effective budget set in estimation.

The results, which are consistent across model specifications, suggest that the labor supply of prime-aged married men is relatively invariant to the net wage and virtual income. The finding of no virtual income effect, however, starkly contrasts with Hausman's result. Furthermore, the estimated net wage elasticities are positive and of larger magnitude than the one reported by Hausman. The results for women are more sensitive to the specification of the labor supply function. Net wage elasticities resulting from a censored estimator are similar to those of Hausman. But when a truncated estimator is used (conditioning on positive hours), estimated wage elasticities are much smaller. The same is also true (in absolute value) of the virtual income elasticities.

Blomquist and Hansson-Brusewitz (1990) estimate a potpourri of labor supply functions for married men and women in Sweden. They consider both linear and quadratic supply functions, with and without random preferences. For males, the linear fixed-preference specification is estimated first with the non-convex budget set and then using the convex hull as an approximation. The random-preference linear specification also uses this convex approximation. The fixed-preference model is a special case of the random-preference model when the constraint $\sigma_\xi^2 = 0$ is imposed. A likelihood-ratio test rejects this constraint at the 1% significance level. The fourth model for males includes a quadratic term in wages. A likelihood-ratio test of the null that this coefficient is zero is rejected at the 1% significance level. All estimation results for males imply a substantial compensated wage rate elasticity and a smaller income elasticity.

For females, the authors correct for sample selection bias using Heckman's two-stage technique. They offer four specifications for female labor supply: (i) linear supply function, fixed preferences, Heckman method; (ii) quadratic supply function, fixed preferences,

Heckman method; (iii) linear supply function, random preferences, Heckman method; (iv) linear supply function, fixed preferences, full-information maximum likelihood. As was the case for men, an asymptotic likelihood-ratio test rejects the null hypothesis of fixed-preferences. The wide differences in compensated wage elasticities between women and men, which are reported in Tables 1 and 2, are somewhat misleading since the wage rate elasticities for both groups are evaluated at different points on the labor supply functions. Using a quadratic supply function and evaluating the female wage rate elasticity at the mean male sample values yields an estimate of 0.10, comparable to the 0.12 estimate for males.

Bourgiugnon and Magnac (1990) estimate labor supply functions separately for a sample of French married men and women, using a piece-wise linear constraint and a convexified budget set. They assume that family labor-supply decisions are sequential, with the men first choosing their labor supply under the assumption of no other labor income in the family. Then the other family members choose their own labor supply, taking the household head's labor supply as given. Under the assumption that (ε,ξ) is distributed bivariate normal $(0,0, \sigma_\varepsilon^2,\sigma_\xi^2,0)$, where ε represents preference heterogeneity and ξ is a measurement error term, the authors estimate the model using maximum likelihood. The authors also consider the joint labor supply model, assuming that the original kinked budget constraint is approximated by some differentiable function as in Section 6.1.3. They use an instrumental variables estimator to estimate this model.

Flood and MaCurdy (1992) apply the full spectrum of methods for convex budget sets to a 1983 cross section of prime-age, married, Swedish men from the Swedish Household Market and Non-market Activities Survey (HUS), in hopes to reconcile the discrepant results of previous work on the disincentive effects of Swedish income taxes. They consider the piece-wise linear and differentiable constraint approaches, estimation using both instrumental variables and maximum likelihood, various functional forms for both labor supply and the structure of measurement error in hours worked, and extensions to incorporate family labor supply and lifecycle considerations. The authors also explore the viability of the standard exogeneity assumptions that underlie the maximum likelihood estimation approach.

Flood and MaCurdy report maximum likelihood results for the following specifications: piecewise-linear and the differentiable method with additive errors, linear labor-supply with and without multiplicative error, and logarithmic labor supply with and without multiplicative error.[46] These specifications yield uncompensated and compensated wage elasticities of around 0.15 and 0.20, slightly higher than those reported by Blomquist (1983).[47] The authors note the minor consequences both of accounting for measurement error and of using the piece-wise linear as opposed to the differentiable approach. This is

[46] See Section 6.2.2 for a discussion of the multiplicative measurement error structure and the logarithmic labor supply function. Also, recall from Section 6.4.3 that specifications relying on differentiable budget constraints need not assume any measurement error to render the empirical model data-consistent.

[47] Blomquist and Newey (1997) find slightly lower wage elasticities and slightly higher income elasticities using their non-parametric formulation of the piece-wise linear labor supply model.

consistent with the findings of Hausman (1981). The instrumental variable estimation results are summarized in Table 2. The key insight from these results is that the data reject the exogeneity assumptions maintained by the maximum-likelihood procedures. These assumptions dramatically influence the estimates of the substitution and income effects; conventional endogeneity tests reject the exogeneity of gross wages and all components of non-labor income. Finally, the results of Flood and MaCurdy suggest that altering the form of the structural labor-supply function produces only small changes, and neither lifecycle adjustments in the computation of virtual income nor attempts to explore the interaction of husband's and wife's labor choices substantively change the results.

Blundell et al. (1998a) present instrumental variable estimates of a labor supply model for the hours of work of married women in the UK that accounts for the endogeneity of gross wages and other income as well as accounting for selection and non-linear taxation. This model and its results are fully documented in our discussion of difference-in-differences specifications in Section 5.

6.8.3. Multiple welfare program participation

We close this section with a look at the labor supply effects of multiple welfare programs, as addressed in the working paper by Keane and Moffitt (1995).[48] They use a single-actor labor supply model to consider the joint decision of whether to work, whether to participate in AFDC, and whether to participate in the Food Stamps program. This necessitates estimation of the labor supply equation jointly with two welfare participation equations to account for the correlation between unobservables. The authors limit agents to full-time, part-time, and no work. Together with the 0/1 decision for two welfare programs, this implies twelve alternatives over which the utility function must be maximized.

Keane and Moffitt estimate the model using a sample from the 1984 SIPP of 968 female heads of households with children. Explanatory variables used include education, age, number of children, region, SMSA, and state characteristics. Using their estimates, they compute the uncompensated wage elasticity, at variable means, as 1.94. This is at the high end of prior estimates, which seems reasonable since this is a study of female-heads rather than married women. They estimate an income elasticity of -0.21, a small (in absolute value) estimate which they attribute to measurement error in unearned income. The estimate of λ, the parameter indicating the extent to which welfare stigma is additive, is 0.05.

In addition, the authors simulate policy changes in the AFDC and Food Stamp programs. First, they consider the impact on predicted choices of reducing the AFDC benefit reduction rate from 100% to 50%. This has limited effect on labor supply but increases both AFDC and Food Stamp participation. Second, they find that a reduction of both AFDC and Food Stamp benefit reduction rates to 10% would increase average labor supply by two hours, but would also increase AFDC participation by one third and Food

[48] See also the discussion of family labor supply and program participation models in Section 7.

Stamp participation by one fourth. This would lead to an 80% increase in net costs even accounting for the increase in tax revenue. Third, they find that increasing gross wages by one dollar would increase average labor supply by about 3.5 h and reduce AFDC and Food Stamp participation, but that a government financed minimum wage of five dollars could accomplish the same changes at lower cost. Finally, they simulate the impact of the 1981 increase in the AFDC tax rate from 67% to 100%, by comparing predictions for the 1984 sample using both the 1980 and the 1984 welfare rules. They find decreased AFDC participation, with many AFDC recipients working part-time in 1980 either leaving AFDC to work full-time or quitting their jobs. As a result, they find an increase of 14.6% in the percentage of AFDC recipients who do not work. All of these results closely match those that were actually observed.

7. Family labor supply

This section considers two important developments to the family labor supply model. The first concerns the extension to cover non-participation and non-convex budget constraints. The second refers to the development of a collective framework for the study of family labor supply. Both are likely to be critical to our understanding of the impact of tax and welfare reforms discussed in Section 2 and our interpretation of the changing patterns of female and male labor supply documented in Section 3.

We develop the analysis of non-participation and non-convex budget constraints in a family labor supply context in two steps. The first simply accounts for non-participation via a corner solution in the labor supply of one of the individuals. The second incorporates a more general specification for welfare programs and fixed costs.

The discussion of the collective labor supply model that follows draws heavily from the recent literature on the specification and identification of these models. We also consider the robustness to alternative model specifications and to the introduction of home production. We round up this section with a review of the results from recent empirical applications of the family labor supply model.

7.1. The basic economic model of family labor supply

The standard approach to family labor supply modeling, discussed in Section 4.1.2, extends the consumption-leisure choice problem to include two leisure decisions. As will be clear from our discussion of collective family labor supply models in Section 7.2, this simple extension of the standard model is controversial. However, it is attractive because it extends naturally to cover multiperiod labor supply decisions[49] and, perhaps more interestingly, it can be used to place the discussion of non-linear budget constraints, fixed costs and participation problems introduced in Section 6 in a family labor supply setting.

[49] In Section 8, we consider in detail the issues that arise in a multiperiod labor supply model with participation.

7.1.1. Family labor supply with participation

The standard family labor supply model concerns the labor supply behavior of a household comprised of two working-age individuals. Children and other dependants are included in the vector of observable household characteristics, X. We assume that families maximize joint utility over consumption, C, and the leisure time of both workers $U(C,L_1,L_2,X)$ where L_1 and L_2 are the hours of leisure for two family members. For expositional reasons, we also consider non-participation for the second individual. The first-order conditions for this problem (see Eqs. (4.15) and (4.16)) can be written

$$U_{L_1} = \lambda W_1 \qquad \text{and} \qquad U_{L_2} \geq \lambda W_2, \tag{7.1}$$

where strict equality holds in the latter marginal condition when individual 2 works. Substituting out for the marginal utility of consumption $U_c = \lambda$ results in

$$U_{L_1} - U_c W_1 = 0 \qquad \text{and} \qquad U_{L_2} - U_c W_2 \geq 0, \tag{7.2}$$

in which each marginal utility is a function of L_1 and L_2 since from the budget constraint we can write consumption as $C = Y + W_1(T - L_1) + W_2(T - L_2)$.

The optimal labor supply choices in this framework satisfy the standard consumer demand restrictions of symmetry, negative semidefiniteness of the Slutsky substitution matrix, and homogeneity of degree zero in wages, prices and full income. Homogeneity is satisfied by specifying the labor supply model in terms of real wages and real incomes. Symmetry requires equality between the Slutsky cross-substitution terms

$$\frac{\partial L_i}{\partial W_j} + L^j \frac{\partial L_i}{\partial M} = \frac{\partial L_j}{\partial W_i} + L^i \frac{\partial L_j}{\partial M} \qquad \text{for } i \neq j. \tag{7.3}$$

The negativity restriction generalizes the Slutsky condition on the sign of compensated labor supply by requiring the matrix of the own- and cross-Slutsky substitution terms to be negative semidefinite. To complete the specification, we may add taste heterogeneity terms to the marginal utility conditions to produce

$$U_{L_1} - U_c W_1 - \varepsilon_1 = 0 \tag{7.4}$$

and

$$U_{L_2} - U_c W_1 - \varepsilon_2 \geq 0, \tag{7.5}$$

with joint density $g(\varepsilon_1,\varepsilon_2)$. These terms are introduced directly into marginal utility rather than into the labor supply equations themselves (in contrast to Section 6) to preserve the taste heterogeneity interpretation of the error terms in a model with multiple labor supply decisions.

These first order conditions describe two regimes of behavior:

(i) both spouses participate: $H_1 \equiv T - L_1 > 0, H_2 \equiv T\text{-}L_2 > 0$,
(ii) individual 2 does not participate: $H_1 \equiv T - L_1 > 0, H_2 \equiv T\text{-}L_2 = 0$,

where H_1 and H_2 are the hours of work choices of each of the two adults in the family.

The sample likelihood for this model has two contributions and is similar to the sample likelihood for the single worker corner solution model described in Section 6. Ignoring taxation and measurement error, and additionally assuming wages are known and exogenous, the likelihood contribution for families observed in the first regime where both spouses work is given by

$$l_{H_1>0,H_2>0} = |J| g(U_{L_1} - U_C W_1, U_{L_2} - U_C W_2), \tag{7.6}$$

where the term $|J|$ is the Jacobian term that corresponds to Eq. (6.15) in the single worker case. This term is the determinant of the own and cross derivative matrix of ε_1 and ε_2 in terms of hours of work. It recognizes that ε_1 and ε_2 are non-linear functions of H_1 and H_2.

For the non-participation regime we note that $\varepsilon_2 > U_{L_2} - U_C W_2$ defines a reservation wage condition, so that the choice of L_1 involves solving the marginal conditions with $L_2 = T$ which we write as $\tilde{U}_{L_1} - \tilde{U}_C W_1 - \varepsilon_1 = 0$. Consequently, the likelihood contribution for observations on families in the regime where the second worker does not participate is given by

$$l_{H_1>0,H_2=0} = |K| \int_{U_{L_2}-U_C W_2}^{\infty} g(\tilde{U}_{L_1} - \tilde{U}_C W_1, \varepsilon_2) \mathrm{d}\varepsilon_2, \tag{7.7}$$

where again the term $|K|$ is the corresponding Jacobian term. It is interesting to note that the Slutsky symmetry and negativity conditions are sufficient to guarantee that both of the matrices J and K in the Jacobian terms are positive definite.[50]

Missing wages, and also the endogeneity of gross wages, is best addressed by rewriting the marginal conditions (7.4) and (7.5) so that they are log linear in wages, i.e.

$$\ln\left(\frac{U_{L_1}}{U_C}\right) - \ln W_1 - \tilde{\varepsilon}_1 = 0 \tag{7.8}$$

and

$$\ln\left(\frac{U_{L_2}}{U_C}\right) - \ln W_2 - \tilde{\varepsilon}_2 \geq 0, \tag{7.9}$$

in which case wage equations of the form $\ln W_j = Z_j' \gamma_j + \xi_j$ can be easily incorporated. Education variables, typically excluded from preferences but included in the Z variables in each wage equation, can then be used to identify the model – under the strong assumption that education is uncorrelated with unobserved heterogeneity in labor supply.

Finally, in order to estimate the wage equation on the sample of observed wages for which $H_2 > 0$, one needs to account for the selection bias induced by correlation in the unobservables and ξ_2. The parameters of the wage equation are identified through the exclusion of the exogenous income variable Y which does enter the determination of participation.

[50] See Ransom (1987) and Van Soest et al. (1990).

7.1.2. Family labor supply with taxes and program participation

The extension of these models to allow for convex piecewise linear budget constraints is a straightforward adaptation of the discussion presented in Section 6. Non-convexities in the budget constraint and welfare program participation pose further difficulties because direct comparisons of utilities are required as documented in Section 6.

Consider the problem of jointly modeling the work and welfare participation decisions of a two-worker family. Suppose we assume that families maximize a standard utility function of the form

$$U = U(L_1, L_2, C, \varepsilon) - \eta P_B, \tag{7.10}$$

where, in keeping with the notation in Section 6.6.3, P_B is a 0-1 program participation indicator. Unobservable preference heterogeneity is entered directly in utility through the vector ε which correspond to the ε_1 and ε_1 terms in Eqs. (7.4) and (7.5). As in Eq. (6.49) the ηP_B term is included so as to capture the costs of being on welfare, including "welfare stigma". The budget constraint that determines consumption is given by

$$C = W_1 H_1 + W_2 H_2 + Y - T(Y, W_1 H_1, W_2 H_2) + BP_B, \tag{7.11}$$

where Y is unearned income, $T(\cdot)$ is a tax function, and B is program benefits.

Due to the computational difficulties encountered when considering the hours and participation of two persons with non-linear budget constraints, the approach outlined in Section 6.7 offers a tractable method for estimating the family labor supply model. In particular, given an assumed joint distribution for unobservable tastes components, errors determining wages, and welfare stigma, one can compute a probability that each family member selects among alternative employment and program participation states. This in turn defines a sample log likelihood of the form Eq. (6.57). As described in Section 6.7, this formulation requires that each individual be placed into a limited set of preassigned work states, even though observed hours worked take many more values. To overcome this issue, analyses applying this approach invariably introduce measurement error in hours of work to admit hours to deviate from the discrete values assumed for the choice set, as described in likelihood (6.58).

7.1.3. Drawbacks of the standard family labor supply

The "unitary" model described in Section 7.1 implies three broad groups of testable restrictions. The first set of restrictions covers the standard consumer demand restrictions of symmetry, negative semidefiniteness of the Slutsky substitution matrix, and homogeneity of degree zero in wages, prices and full income, see Eq. (7.3) and the related discussion above. The second set of restrictions refer to income pooling. This is the condition which implies that, as far as the household's utility-maximizing choice of family labor supplies are concerned, one can combine all sources of non-labor income into a single unearned income measure, Y. If, for example, each of the two individuals has *private* unearned income Y_1 or Y_2 respectively, then pooling implies

$$L_i^{Y_1} = L_i^{Y_2}, \qquad \text{for } i = 1 \text{ and } 2, \text{ where } L_i^{Y_1} \equiv \frac{\partial L_i}{\partial Y_1}. \tag{7.12}$$

This is a controversial assumption in the welfare reform debate since it implies that the source of non-labor income is irrelevant in within-family labor supply decisions.

Finally, there are the non-participation or "corner solution" conditions which state that if one individual is at a corner solution, it is the reservation wage of that individual rather than the market wage that affects the labor supply decision of the partner. As in the case of the income pooling assumption, this is far from innocuous, implying as it does that the "outside option" value of paid work for a non-participant does not influence the allocation of consumption and leisure within the household.

7.2. The collective model of family labor supply

Recent research has focused on relaxing the assumptions of symmetry and income pooling, seeking instead solutions from efficient bargaining theory. The advantages of the unitary model of family labor supply are well known. As we have seen they allow the direct utilization of consumer theory, recovering preferences from observed behavior in an unambiguous way and providing a framework for interpretation of empirical results. One can then use standard welfare economics to evaluate tax and welfare reform. An argument often raised by critics of the standard model is that it treats all individuals in the family as a single decision making unit rather than as if they were a collection of individuals. Moreover, researchers often conclude that allocations within the family derived from the unitary model cannot be recovered in a meaningful way. This conclusion is too strong. The standard decentralization theorems from consumer theory[51] apply equally well to individual members' utilities in a "unitary" household.

Suppose there are no public goods and that individual utilities are weakly separable over their private consumption and leisure. Let C_1, C_2, L_1, and L_2 refer to the private consumption and leisure choices of individuals 1and 2. Defining the private consumption of the second individual in the same way, we may write the within-period family utility as

$$F(U_1(C_1, L_1, X), U_2(C_2, L_2, X)), \tag{7.13}$$

where $U_1(C_1,L_1,X)$ is the sub-utility for the husband and $U_2(C_2,L_2,X)$ is the sub-utility for the wife. Where family utility has this weakly separable form, decentralization follows two-stage budgeting: total household (full) income is allocated among all household members, and then individuals act as if they are making their labor supply and consumption decisions conditional on this initial stage outlay.

Even if consumption goods are privately consumed, they are typically only measured at the household level – so that the individual consumptions are unobserved or "latent" to the economist. However, a single observed (privately-consumed) good – labor supply in this case – per sub-utility is often sufficient to identify decentralized preferences. This condi-

[51] See Gorman (1958), for example.

tion on a single exclusive good per sub-utility corresponds to the identification condition in generalizations of weak separability that allow overlapping goods across groups.

So what advantages does the collective approach offer? It effectively relaxes the income allocation rule among individuals so that this allocation may depend on relative wages and other variables in a way that reflects the bargaining position of individuals within the family, rather than reflecting the marginal conditions underlying the joint optimizing framework of the traditional unitary approach. Even when individuals within the family are altruistic and allocations are Pareto Efficient, the allocation rule can deviate from the optimal rule in the traditional model.

7.2.1. A summary of the collective labor supply model model

In this work,[52] each family member either maximizes an "egoistic" utility, $U_1(C,L_1,X)$, or a "caring" utility function, $F_j(U_1(C_1,L_1,X), U_2(C_2,L_2,X))$, for $j = 1$ and 2. Notice that this mirrors the separability assumption in Eq. (7.13). That is, the only way L_2 enters the (sub-) utility of individual 1 is through the (sub-)utility of individual 2; there is no direct impact on the utility of the partner.

Applications of this model assume that the decision process generates Pareto-efficient outcomes, all goods are privately consumed and there is no household production. The implications of relaxing these latter two assumptions are important and we consider them below.

The collective framework states the family labor supply problem as follows:

$$\max[\theta U_1 + (1-\theta)U_2], \qquad \text{s.t.} \qquad C_1 + C_2 + W_1L_1 + W_2L_2 = M, \tag{7.14}$$

where θ is the utility weight for person 1, given by some non-negative function $\theta = f(W_1, W_2, M)$. This is equivalent to a sharing rule, or decentralized solution, in which individual 1 gets income $M - \varphi(W_1, W_2, X, M)$ and then allocates according to the rule

$$\max U_1, \qquad \text{s.t.} \qquad C_1 + W_1L_1 = M - \varphi(W_1, W_2, X, M), \tag{7.15}$$

where $\varphi(W_1, W_2, X, M)$ is defined as the sharing rule.

Given Pareto efficiency and the standard neoclassical assumptions on individual utilities, the conditions identifying preferences and the sharing rule (up to a linear translation) simply require one observable and assignable private good – here assumed to be the individual's leisure. The intuition behind identification is simple: under the exclusive good assumption, the spouse's wage can only have an effect through the sharing rule. Variation of income and the wage then permit consistent estimation of the marginal rate of substitution in the sharing rule. A researcher can do this for both spouses and, since the sharing rule must sum to one, recover the partial derivatives of the sharing rule.

Although the standard symmetry, income pooling, and participation conditions are not implications of this model, one can derive alternative testable restrictions. If separate

[52] The most lucid statement of this argument occurs in the papers on household labor supply by Chiappori (1988, 1992).

income sources are unobservable to the econometrician and both individuals work in the labor market (i.e., there are no corner solutions for leisure), the only restrictions implied are those corresponding to the Slutsky conditions. These are expressed in terms of the derivatives of the labor supply equations with respect to the wage and income variables. Assuming the income derivatives are non-zero the collective model implies the differential equations:

$$\alpha_m \frac{L_1^{W_2}}{L_1^M} + \alpha \frac{\partial}{\partial M} \frac{L_1^{W_2}}{L_1^M} - \alpha_{W_2} = 0, \qquad \beta_M \frac{L_2^{W_1}}{L_2^M} + \beta \frac{\partial}{\partial M} \frac{L_2^{W_1}}{L_2^M} - \beta_{W_1} = 0, \tag{7.16}$$

in which α is given by

$$\alpha = -\left[\frac{\dfrac{\partial}{\partial M}\left(\dfrac{L_1^{W_2}}{L_1^M}\right)\dfrac{L_2^{W_1}}{L_2^M} - \dfrac{\partial}{\partial W_2}\left(\dfrac{L_2^{W_1}}{L_2^M}\right)}{\dfrac{\partial}{\partial M}\left(\dfrac{L_2^{W_1}}{L_2^M}\right)\dfrac{L_1^{W_2}}{L_1^M} - \dfrac{\partial}{\partial W_1}\left(\dfrac{L_1^{W_2}}{L_1^M}\right)} \right]^{-1},$$

$\beta = 1 - \alpha$ and where superscripts denote partial derivatives. The terms $\alpha_M, \beta_M, \alpha_{wj}, \beta_{wj}$ are the corresponding income and wage derivatives. Eqs. (7.16) are analogous to the Slutsky symmetry conditions, while the Slutsky inequalities are matched by

$$\frac{L_1^{W_1}}{L_1^M} + \left(T - L_1 - \frac{\beta}{\alpha}\frac{L_1^{W_2}}{L_1^M}\right) \leq 0, \qquad \frac{L_2^{W_2}}{L_2^M} + \left(T - L_2 - \frac{\beta}{\alpha}\frac{L_2^{W_1}}{L_2^M}\right) \leq 0. \tag{7.17}$$

These restrictions are sufficient for recovering preferences and the sharing rule (up to an additive constant). Indeed, the derivatives of the sharing rule, $\varphi(W_1, W_2, X, M)$, have the form

$$\frac{\partial \varphi}{\partial M} = \alpha, \qquad \frac{\partial \varphi}{\partial W_2} = \frac{\partial L_1^{W_2}}{\partial L_1^M}\alpha, \qquad \frac{\partial \varphi}{\partial W_1} = \frac{\partial L_2^{W_1}}{L_2^M}(\alpha - 1). \tag{7.18}$$

Consequently, having estimated unrestricted family labor supply functions in terms of wages for each individual and full income, the researcher can recover individual preferences and the sharing rule.

7.2.2. *Household production*

The introduction of household production is problematic for estimation of the collective model since, as we have seen, this model exploits the exclusion restriction on the other individual's wage to identify the sharing rule under egoistic or caring preferences. Unless we assume that the household production good is marketable, identification up to an additive constant is lost.

It is reasonable to assume that for many families, non-market time is spent in the active production of home produced goods.[53] These may include activities for which a perfect substitute is directly available in the market, housework or home decoration for example;

[53] See also Apps and Rees (1997, 1998).

but they may also include activities for which a perfect substitute is not readily available, childcare, for example. What is of particular interest is when there is no direct substitute available and both spouses non-market time enter the production of the home produced good. For the standard non-separable (unitary) model of household labor supply, this has little direct impact on the labor supply function – it simply acts as if it were leisure time. However, that is not the case in the separable model.

Suppose there is a home-produced good, G, that requires inputs of time by both household members. Denoting these time inputs by t_1 and t_2, one can write the production technology as

$$G = g(t_1, t_2), \tag{7.19}$$

where we assume that g is a concave function. Time not spent in the labor market can be used for two purposes, pure leisure or home production. If t_1 and t_2 are recorded by individuals in a time-use diary survey,[54] then the characteristics of $g(t_1, t_2)$ can be recovered. However, since leisure enters household utility in a general way in the family utility function, and since $g(t_1, t_2)$ is concave, family utility remains a concave function of non-market time and consumption. Consequently, the labor supply equations describing hours of work and labor market participation are observationally equivalent to those for the model without household production.

An interesting special case occurs when family utility is separable in the non-market time of each individual. In this case, family utility with household production can be written

$$F(U_1(C, L_1, G_1, X), U_2(C, L_2, G_2, X)), \tag{7.20}$$

where G_1 and G_2 are the private consumptions of the home-produced good so that $G_1 + G_2 = G$. (Alternatively, if the home-produced good is a public good such as childcare, then G itself enters each sub-utility.)

If the consumption of household production is not observed then the presence of G in each sub-utility does upset the separability assumption. To see this, suppose household production technology exhibits constant returns to scale. Then the implicit price, or unit cost, of household production is simply a function of the two wage rates:

$$P^* = r(W_1, W_2). \tag{7.21}$$

In the model without G, the weak separability condition is sufficient for each labor supply to be written in terms of the own wage and the allocation of full income. Introducing G implies that P^*, and therefore W_1 and W_2, enter each labor supply. Consequently, the household production function is sufficient to break the separability condition and therefore the exclusion restriction on the other household member's wage in the labor supply equation. In this case, individual utilities are not recoverable. The only case in which this does not occur is when the household production good, G, is marketable and when the

[54] See Kapteyn and Kooreman (1993), for example.

solution is interior rather than at a corner. In this case, one sets P^* equal to the observable market price for the home produced good which will not depend on individual wages. This issue becomes more problematic for the collective model described below in which the exclusion restriction on the other individual's wage is required for identification. Of course, if the household production technology exhibits constant returns then P^* in (7.21) depends only on the two wages and the income terms in the sharing rule provide testable restrictions.[55]

7.3. Some empirical findings for the family labor supply model

7.3.1. The unitary model

Recent studies build on the original work of Ashenfelter and Heckman (1974), Rosen (1978), Wales and Woodland (1976) and Smith (1977). These include Attanasio and MaCurdy (1997), Blundell and Walker (1982, 1986), Browning et al. (1985), Hausman and Ruud (1986), Hoynes(1996), Kooreman and Kapteyn (1986), Ransom (1987) and Van Soest (1995). Perhaps the most important issues raised in these studies are those concerning the degree of substitution between the labor supplied by different family members and the mixture of continuous hours and discrete participation choices.

Researchers have taken two modeling approaches. The first is to work with the bivariate censored model and allow continuous choices over hours of work. Ransom (1987), for example, takes this approach. The second approach is to simplify the hours choices to a set of discrete alternatives but to allow for fixed costs and program participation. Hoynes (1996) is an example of this. In addition to accounting for the discrete or censored nature of the data in a bivariate framework, researchers who have implemented empirical models of family labor supply have also been concerned with choosing the appropriate conditioning variables. Attanasio and MaCurdy (1997), for example, adopt a marginal rate of substitution framework for their analysis while Blundell and Walker (1986) use a consumption-based measure of non-labor income in a Marshallian model of family labor supply. Browning et al. (1985) work with a Frisch representation of family labor supplies and commodity demands (see Section 4 for a detailed discussion of these alternative choices of conditioning variables and the interpretation of the resulting elasticities). Rather than covering all studies in this discussion we have decided to single out a small number of studies that provide a useful guide to empirical models in the literature.

7.3.2. Continuous hours models with censoring

Ransom (1987) provides an analysis of family hours-of-work decisions using a sample of 1210 intact families drawn from the 1976 PSID. He restricts the sample to families with no self-employed members and in which the husband is working. Consequently, the only censoring occurs for female hours of work. This study makes a particularly convenient starting point for describing structural estimation in family labor supply models. However,

[55] Chiappori (1997) shows identification of the sharing rule up to some function of W_1 and W_2 in this case.

although Ransom accounts for censoring he does not account for the endogeneity of hourly wages or virtual income. Attanasio and MaCurdy (1997) relax these exogeneity restrictions and we discuss their study further below.

To interpret Ransom's results, consider the marginal utility conditions (7.4) and (7.5) for quadratic utility:

$$-UL_1 + W_1U_C = \alpha_1 + \alpha_3W_1 - \beta_{11}H_1 - \beta_{33}W_1(W_1H_1 + W_2H_2 + Y) - \beta_{12}H_2 + \beta_{13}(2W_1H_1 + W_2H_2 + Y) + \beta_{23}W_1H_2 \tag{7.22}$$

and

$$-UL_2 + W_1U_C = \alpha_2 + \alpha_3W_2 - \beta_{22}H_2 - \beta_{33}W_2(W_1H_1 + W_2H_2 + Y) - \beta_{12}H_2 + \beta_{23}(2W_2H_2 + W_1H_1 + Y) + \beta_{13}W_2H_1. \tag{7.23}$$

Ransom then allows α_1 and α_2 to each be a linear function of observable characteristics and unobservable mean zero normal random variates ε_1 and ε_2.

In contrast, Hausman and Ruud (1986) work directly with quadratic labor supply curves in their estimation of family labor supply for a sample of 1991 families in the 1976 PSID based on a similar selection to that of Ransom. This eliminates the need for jacobian terms but makes the introduction of random preference errors more difficult. The Hausman and Ruud indirect utility has the form

$$V(W_1, W_2, Y) = exp(\beta_1W_1 + \beta_2W_2)Y^*, \tag{7.24}$$

where Y^* is given by the quadratic

$$Y^* = Y + \theta + \delta_1W_1 + \delta_2W_2 + 0.5(\gamma_1W_1^2 + \gamma_2W_2^2 + \alpha W_1W_2. \tag{7.25}$$

From Roy's identity, hours of work are given by

$$H_1 = \delta_1 + \beta_1Y^* + \gamma_1W_1 + \frac{\alpha}{2}W_2, \tag{7.26}$$

$$H_2 = \delta_2 + \beta_2Y^* + \gamma_2W_2 + \frac{\alpha}{2}W_1, \tag{7.27}$$

where Y^* is defined in Eq. (7.25). Hausman and Ruud append additive normal errors to Eqs. (7.26) and (7.27) and estimation follows from the maximizing of a bivariate censored likelihood whose contributions are of the form (7.6) and (7.7). They take careful account of the non-linear surface of the budget constraint induced by the piecewise linear nature of the tax system. However, as was noted above, one cannot easily interpret this additive error structure as random preference variation.

The estimates Ransom presents are plausible. He finds a small compensated elasticity of 0.04 for men, and a larger one of 0.73 for women. He also finds smaller income elasticities for men, −0.03 versus −0.21 for women. (The uncompensated wage elasticity for men is

slightly negative.) These average elasticity estimates are close to those reported in Hausman and Ruud. However, the Ransom model does not perform particularly well in replicating the within-sample distribution of female hours of work. While actual mean hours of work for women are 1376 per year, the model predicts only 634. Moreover, although fewer than 50% of women in the sample are recorded as working, the model predicts that the majority of women work. Although there are many ways in which the model might be misspecified, the most likely culprits are the "Tobit" assumption on participation that rules out fixed costs of work (see Cogan, 1980, 1981) and the use of predicted wages (although corrected for selectivity) in a non-linear labor supply model. Interestingly, the Hausman and Ruud specification allows for a fixed cost parameter for female labor supply which is found to be significant.

Attanasio and MaCurdy (1997) address endogeneity issues and generalize the form of censoring in their study of families in the repeated cross-sections of the US Consumers Expenditure Survey (CEX). They choose log-linear forms for the marginal rate of substitution functions Eqs. (7.8) and (7.9). The authors selected a sample of 20,297 households from the CEX for the period 1981–1992. This CEX dataset has the dual advantages of providing consumption data directly and allowing variation over time. Attanasio and MaCurdy adopt a semi-parametric approach to correct for non-participation, and this relaxes the normality and Tobit assumptions. Their results imply a slightly negative male hours elasticity, whereas for women the corresponding elasticity is much lager in absolute value and implies a strongly upward sloping labor supply curve.

Kooreman and Kapteyn (1986) follow a similar approach but do not fully allow for random preference variation, in their study of the joint labor supply decisions of 315 households from the 1982 CBS survey for the Netherlands. As in Blundell and Walker (1986) they work directly with a Marshallian demand system and specify a second-order flexible form, derived from an Almost Ideal indirect utility function. In contrast, the Blundell and Walker study uses a Gorman polar form which retains the linearity in full income. This linearity obviates the need to specify a value for total available hours T. Kooreman and Kapteyn perform estimation using a censored likelihood in which they predict wages from a selectivity-adjusted log wage equation. They use education-level dummies as excluded instruments. Although Kooreman and Kapteyn do not provide elasticity estimates, they do report small and negative own wage responses for men and larger and positive own wage responses for women. Cross elasticities show a strong response of female hours to male wages. These results closely match those of Blundell and Walker who used a truncated likelihood estimator on a similar sample of 1378 British families from the 1980 Family Expenditure Survey, in which both husbands and wives worked. Blundell and Walker report a full set of elasticities by demographic type, and find small, positive labor supply elasticities for males and larger, positive ones for females. They also report small but significant positive cross elasticities.

7.3.3. Discrete hours choices and program participation

Van Soest (1995) introduces discrete hours choices in a study of the labor supply and

participation decisions of a sample of 2859 families from the 1987 Social Economic Panel for the Netherlands. He models a non-convex budget constraint for each family to explicitly account for the Dutch tax and benefit system. Six fixed hours intervals are defined, for husbands and for wives, resulting in a total of thirty-six possible discrete states.[56] A translog direct utility function for leisure hours and full income determines the utilities associated with each choice. Marginal utilities are, therefore, linear in log leisure hours and full income and are rendered stochastic by a choice-specific extreme-value distributed error term. Consequently choices follow a multinomial logit rule. Van Soest further extends this by introducing a jointly normal random parameter variation. He completes the specification by adding choice-specific constants and estimates using a simulated maximum likelihood estimator.

The estimation results reveal an important role for the random preference terms and the choice specific constants. Moreover, the reported elasticities are quite sensitive to changes in their specification. The most general specification suggests a small positive hours of work elasticity of around 0.1 for men and a larger elasticity of around 0.5 for females, with small, negative cross elasticities. These are reasonably consistent with the results in the Blundell and Walker study for the UK. However, Van Soest does report smaller income elasticities which often have signs at odds with theory. This may be attributable to his unearned income measure which, unlike that used in the Blundell and Walker study, is not consumption based.

Hoynes (1996) addresses the problem of jointly modeling the discrete work and welfare participation decisions of a two-worker family, in the context of the unitary labor supply model for a sample of 1010 observations on two-parent families from the1984 Survey of Income and Program Participation (SIPP). She considers the labor supply impacts of the AFDC-UP (Aid for Families with Dependent Children – Unemployed Parent) program in the US in 1988. This program was available in 26 states in 1988 and provided AFDC benefits to two-parent families with children if the "principal earner" in the family worked less than one hundred hours/month. Hoynes models families as maximizing a standard Stone–Geary utility function (see Eq. (4.59)) of the form

$$U = \beta_1 \log(\gamma_1 - H_1) + \beta_2 \log(\gamma_2 - H_2) + \beta_C \log(C - \gamma_C) - \eta P_B, \tag{7.28}$$

where the notation is as in Eq. (7.10) and P_B is a 0-1 program participation indicator. Program benefits B in (7.11) are determined by

$$B = G - N - t_A(W_1 H_1 + W_2 H_2), \qquad \text{if } B > 0 \text{ and } H_p < 100, \tag{7.29}$$

with $B = 0$ if the conditions are not met, where G is a government minimum guarantee, t_A is a government set benefit-reduction rate on earned income, and H_p is hours worked by the principal earner, either the husband or the wife.[57] Hoynes assumes that families who

[56] See also Ilmakunnas and Pudney (1990), Dickens and Lundberg (1993) and Aaberge et al. (1995) for important variations on this type of model that allows finite discrete choice sets.

[57] The principal earner is determined by program guidelines.

choose to receive AFDC-UP ($P_B = 1$) also receive food stamps. To avoid the issue of multiple program choices, (see the discussion of the Keane and Moffitt study in Section 6) and to make the problem manageable, she also assumes that only three work decisions are possible for each spouse: full-time work (40 h per week), part-time work (20 h per week), or no work. Combined with a 0/1 welfare decision, this yields 18 states over which the utility function must be maximized.

Hoynes introduces unobserved heterogeneity into the problem via the β and η parameters. She models the β's as

$$\beta_i = \exp(X'\alpha_i + \varepsilon_i)/\Sigma[\exp(X'\alpha_i + \varepsilon_i)], \qquad i = 1, 2, c, \tag{7.30}$$

where $\alpha_c = \varepsilon_c \equiv 0$.[58] η is given by

$$\eta = Z'\alpha_s + \mu + e, \tag{7.31}$$

where $e \sim N(0, \sigma_s^2)$. To further ease the estimation problem she assumes that ε and μ have a discrete support over M points (where $M = 6$ in the analysis) such that

$$\Pr(\varepsilon_1 = \varepsilon_{1k}, \varepsilon_2 = \varepsilon_{2k}, \mu = \mu_k) = \pi_k. \tag{7.32}$$

To understand the computational issues involved, consider the implications of this setup when $e = 0$ for all families. Then, for each of the M points of support, the values of ε and μ can be plugged directly into the utility function and the optimal choice over the 18 alternatives computed. Summing across the M states using the probability, π_k, of each yields the probability of each work/welfare alternative for each family. Replacing a non-zero e term complicates this only slightly – e behaves like the continuous error in a standard discrete choice model. For high enough e in each state a non-welfare option will be chosen, whereas a low e implies choice of a welfare option. Thus, within each state (ε, μ pair), two possibilities exist, with their probabilities determined by the level of ν required to make "welfare stigma" too high for welfare participation. Hoynes uses predicted wages from separate wage regression for all observations. Hence, the variables she uses to predict wages are included in the variables X, Z.

Hoynes assumes a measurement error term for each spouse, ν_1 and ν_2, such that actual hours worked, h_1 and h_2, relate to predicted hours worked, H_1 and H_2, according to the functions:

$$h_1 = \exp(\nu_1)H_1 \qquad \text{and} \qquad h_2 = \exp(\nu_2)H_2, \tag{7.33}$$

where $\nu_j \sim N(-\sigma_j^2/2, \sigma_j^2)$.

Given the parameter estimates from this estimation procedure, Hoynes carries out several interesting simulations. First, she considers the impact of increasing the AFDC-UP guarantee amount, G, by 20%. This leads to an 18% increase in predicted participation in the program. In the population as a whole it leads to a slight reduction in employment, as

[58] This form satisfies the standard Stone–Geary restrictions that $\beta_i \geq 0$ and $\sum \beta_i = 1$.

should be expected since this is a pure income effect. However, among welfare recipients it leads to an increase in average hours worked, a composition effect caused by the addition of many working families whose incomes qualify for the program only after the policy shift. If unaccounted for, such a composition effect could lead researchers to reach incorrect conclusions concerning the impact of guarantees on labor supply. Second, Hoynes considers the impact of lowering the benefit reduction rate, t_A, by 20%. This leads to a 6% increase in participation and virtually no change in employment, as a tax rate change has both income and substitution effects. Third, she considers the impact of eliminating the $h_P < 100$ rule for eligibility. While this greatly increases eligibility, from 10.9% to 15% of the sample, it has almost no effect on program participation, since those who become eligible are already working and, thus, would receive small benefits which may be overwhelmed by other welfare costs (η). Finally, she finds that eliminating the program altogether would increase average hours worked of current recipients by 33 h for women and 46 h for men. However, this would not compensate for the loss of welfare income, as average family income for this group would still fall by approximately \$83/month.

7.3.4. Bargaining and collective models

There are relatively few empirical studies of family labor supply outside the unitary model. The original motivation for these developments came from the original studies by McElroy (1981) and Manser and Brown (1980). A number of more recent studies have used micro data to evaluate the pooling hypothesis or to recover collective preferences using exclusive goods, but these studies typically look at private consumption rather than labor supply. For example, Browning et al. (1996) use Canadian household expenditure data to examine the pooling hypothesis and to recover the derivatives of the sharing rule. Clothing in this analysis is the exclusive good providing identification.

Recent empirical studies concerning family labor supply include Lundberg (1988), Apps and Rees (1997), Kapteyn and Kooreman (1990) and Fortin and Lacroix (1997). Each of these aims to "test" the unitary model and to recover some parameters of collective preferences. Lundberg attempts to see which types of households, distinguished by demographic composition, come close to satisfying the hypotheses implied by the unitary model. The other three studies take this a step further by directly specifying and estimating labor supply equations from a collective specification. Apps and Rees (1996) specify a model to account for household production. Kooreman and Kapteyn (1990) use data on preferred hours of work to separately identify individual from collective preferences and, consequently, to identify the utility weight. Fortin and Lacroix (1997) follow closely the Chiappori framework and allow the utility weight to be a function of individual wages and unearned incomes. We briefly consider the results from each of these studies.

Kooreman and Kapteyn (1990) specify a Stone–Geary model of individual private utilities and they estimate the utility weight, which they assume to be independent of wages and income, as a constant parameter. Using data from the same 1982 Dutch survey

exploited in their 1986 study described above, they find an estimated utility weight within the unit interval, but rather imprecisely determined.[59]

The focus of the Apps and Rees study is on household production and they analyze a sample of 1384 families from the Australian Bureau of Statistics 1985/86 Income Distribution Survey Sample file. All families are selected so that the male works and there is at least one child aged under 15 years. They specify a constant returns technology for household production so that the unit cost function has the form (7.21). This is then parameterized as a unit Translog function. Individual sub-utilities are given an Almost Ideal form. Since the sample does not contain information on individual consumptions of home produced or market goods, they identify the model by setting the individual income shares to the individual full incomes $W_{it} + M_t$. This would appear to be a rather restrictive assumption. Finally, only interior solutions are considered. They find an important role for exchange within the family with the female specializing in home production activity.

Fortin and Lacroix (1997) consider a sample of 4496 couples drawn from the 1986 Canadian Census. They follow the Chiappori framework closely and allow the utility weight to be a function of individual wages and unearned incomes. They specify the resulting sharing rule as a linear function of wages and individual unearned incomes, while they allow indirect utilities to be quadratic in own wages and individual unearned income allocations. For comparison, they specify a unitary model with a quadratic indirect utility in the two wages and total unearned income. Both specifications result in non-linear labor supply equations. For estimation, they use the sample of two working couples with instrumental variable procedures applied for the wage and income variables. Instruments were age and education polynomials, immigration dummies and regional dummies.

Fortin and Lacroix provide results for two age subgroups. For the majority of groups they reject the pooling hypothesis. The collective model restrictions are only rejected for the case in which preschool children are present, while symmetry is rejected across all groups. These results are interesting and, if confirmed across specifications accounting for endogenous participation in work and unobserved heterogeneity, they would challenge the standard family labor supply model. The results also suggest extensions to the collective model for families with young children where "home production" and public goods are likely to be of central importance.

One potentially important drawback of these models is their inability to allow for both preference heterogeneity and non-participation. This is common in modern specifications of the unitary family labor supply model as we have seen in the earlier discussion of this section. To properly assess the collective framework as an alternative empirical model, these developments are essential. This is the motivation for the Blundell et al. (1998a) study which considers the full non-parametric identification of the collective model with participation and hours choices. A general identification result is presented which is then extended to cover the introduction of unobserved heterogeneity. For the heterogeneous

[59] It should be noted that the estimated parameters and their identification rest heavily on their interpretation of data from the preferred hours question in the Dutch survey.

case a parametric form for preferences and the sharing rule is assumed. This result allows the empirical implementation of the collective model of family labor supply to be placed on an equal footing with the traditional model.

8. Structural dynamic models

This section explores extensions of the standard multiperiod model, introduced in Section 4, to allow for important dynamic features of labor supply behavior. The first considers the problem of participation, which plays a fundamental role in understanding all aspects of lifecycle behavior. Empirical models incorporating participation are obviously important for the analysis of female labor supply and retirement decisions. However, even in the simple case of continuous hours decisions examined in Section 4, we could not specify relations useful for policy simulations without assuming when a person works during the lifetime, for specifications depend on past, current, and future wages. If a person plans not to work in a period, then the wage for that period does not enter as a determinant of hours-of-work choices in other periods. To characterize the factors governing when individuals work significantly complicates empirical multiperiod models of labor supply, and the use of these models in simulations of policy scenarios. However, development of these more-elaborate models is essential to learn what is needed to account for many policy features. Given the scarcity of research on this topic, intertemporal models with non-participation or corners and saving offers many research opportunities.

The second extension considers two lifecycle models in which individuals can affect their wage growth through current investment activities: learning-by-doing models in which current work experience enters directly into the determination of future wages, and conventional human capital models in which workers endogenously choose schooling and training separately from work experience to enhance their future wages. Both of these developments imply that future events enter the optimal decision rule for hours of work and participation decisions in a more complex way.

Finally, the third extension relaxes the intertemporal separability assumption on preferences underlying the standard labor supply framework, implying that past levels of hours and consumption directly impact the marginal utility of work. Non-separabilities occur through primarily two routes: a habit persistence model, or a dynamic extension of the home production model in which inputs of time are used to produce future consumption.

8.1. The standard intertemporal labor supply model with participation

This section begins with an overview of an intertemporal labor supply model with participation which will serve as a framework for discussing the additional dynamic refinements in later subsections. Although decisions over continuous hours choices and consumption retain the simple marginal rate of substitution and Euler condition formulation described in Section 4, the participation no longer fits this simple framework. To

highlight the complexities introduced by participation, in this basic multiperiod model we presume that individuals can only choose between working and not working in a period.

8.1.1. Economic formulation

The optimization problem for participation with borrowing and saving is the solution to

$$\max_{P_t} V_t(P_t, A_t, W_t, Z_t), \tag{8.1}$$

where P_t is a zero-one dummy variable equaling unity if the individual participates in period t, V_t is the period-t value function, A_t represents beginning-period assets, W_t denotes period-t earnings from participation, and Z_t designates all non-wage variables relevant for lifecycle decision making that are not controlled by the decision maker. The elements of Z_t may be stochastic, with some uncertain in the future to the consumer. Decisions over time are linked through the asset accumulation constraint

$$A_{t+1} = (1 + r_t)(A_t - C_t + W_t P_t + Y_t), \tag{8.2}$$

where r_t is the return on assets, and Y_t is a component of Z_t representing income not attributable to earning or returns on assets. Eq. (8.2) assumes perfect capital markets.

The formulation for the value function follows from first principles in dynamic economics. Let $U(P_t,C_t,Z_t)$ be the utility function for period t, which need not depend on all or any elements of Z_t; we include Z_t as an argument, rather than some subset of this vector, to save notation. We can write the value function as

$$V_t(P_t, A_t, W_t, Z_t) \equiv V_t^P = P_t V_t^1 + (1 - P_t)V_t^0 = V_t^P(A_t, W_t, Z_t), \tag{8.3}$$

where

$$V_t^1 = \max_{C_t}\left[U(1, C_t, Z_t) + \kappa E_t\left(\max_{P_{t+1}} V_{t+1}^P((1 + r)(A_t - C_t + W_t + Y_t), W_{t+1}, Z_{t+1})\right)\right], \tag{8.4}$$

$$V_t^0 = \max_{C_t}\left[U(0, C_t, Z_t) + \kappa E_t\left(\max_{P_{t+1}} V_{t+1}^P((1 + r)(A_t - C_t + Y_t), W_{t+1}, Z_{t+1})\right)\right],$$

with the operators E_t designating the consumer's expectation about the variables W_{t+1} and Z_{t+1} conditional on information I_t at time t, which includes W_t and Z_t. The term κ is a discount rate. The first-order condition of (8.4) with respect to C_t yields the Euler condition (4.28), which continues to relate the marginal utilities of consumption in adjacent periods even in this model with participation.

Alternative useful expressions for V_t^1 and V_t^0 are

$$V_t^1 = \max_{C_t}[U(1, C_t, Z_t) + \kappa E_t(\text{Prob}(P_{t+1} = 1 \mid I_t)V_{t+1}^1 + \text{Prob}(P_{t+1} = 0 \mid I_t)V_{t+1}^0)], \tag{8.5}$$

$$V_t^0 = \max_{C_t}[U(0, C_t, Z_t) + \kappa E_t(\text{Prob}(P_{t+1} = 1 \mid I_t)V_{t+1}^1 + \text{Prob}(P_{t+1} = 0 \mid I_t)V_{t+1}^0)],$$

where, for instance, $\text{Prob}(P_{t+1} = 1 \mid I_t)$ designates the consumer's probability of making the decision $P_{t+1} = 1$ conditional on information I_t. The value function in the last period, τ, is

$$V_\tau^P = P_\tau V_\tau^1 + (1 - P_\tau)V_\tau^0 = V_\tau^P(A_\tau, W_\tau, Z_\tau), \tag{8.6}$$

where

$$V_\tau^1 = \max_{C_\tau} U(1, C_\tau, Z_\tau) \qquad \text{s.t.} \qquad C_\tau = A_\tau + Y_\tau,$$

$$V_\tau^0 = \max_{C_\tau} U(0, C_\tau, Z_\tau) \qquad \text{s.t.} \qquad C_\tau = A_\tau + Y_\tau, \tag{8.7}$$

Solving recursively using backward induction yields formulations for each period's value functions and optimal choices.

8.1.2. Empirical formulation

An empirical model characterizes how the values of $P_1, P_2, \ldots, P_\tau$ vary across a population, relating these participation decisions to economic factors relevant in the past, now, or in the future. Creating the likelihood function for the P_t's requires specifying the densities describing the joint distributions of the W_t's and Z_t's, and identifying the partitions of W_t's and Z_t's associated with making particular decisions.

Consider, first, decisions in the final period. Define the sets:

$$\Theta_{\tau 1} = \{(W_\tau, Z_\tau) : V_\tau^1 > V_\tau^0\},$$

$$\Theta_{\tau 0} = \{(W_\tau, Z_\tau) : V_\tau^1 \leq V_\tau^0\}. \tag{8.8}$$

For combinations of W_τ and Z_τ falling in the set $\Theta_{\tau 1}$, the individual chooses $P_\tau = 1$; and when $(W_\tau, Z_\tau) \in \Theta_{\tau 0}$ this person does not work in period T. The sets $\Theta_{\tau 1}$ and $\Theta_{\tau 0}$ are functions of all decisions and variables observed in previous periods.

Now considering period $\tau - 1$, define the sets:

$$\Theta_{(\tau-1)1}^1 = \{(W_{\tau-1}, Z_{\tau-1}) : V_{\tau-1}^1 > V_{\tau-1}^0\},$$

$$\Theta_{(\tau-1)}^0 = \{(W_{\tau-1}, Z_{\tau-1}) : V_{\tau-1}^1 \leq V_{\tau-1}^0\}, \tag{8.9}$$

The individual works when $(W_{\tau-1}, Z_{\tau-1}) \in \Theta_{(\tau-1)1}$, and does not work otherwise. Once again, the sets $\Theta_{(\tau-1)1}$ and $\Theta_{(\tau-1)0}$ depend on decisions and variables observed in periods $\tau - 2, \tau - 3, \ldots, 1$.

Letting $g(\cdot)$ denote the joint density function of the W_t's and Z_t's, the probability of the event $(P_1, P_2, \ldots, P_\tau)$ is

$$l_{P_1 P_2 \cdots P_\tau} = \int_{\theta_1 P_1} \cdots \int_{\theta_t P_t} \cdots \int_{\theta_\tau P_\tau} g(W_1, Z_1, \ldots, W_\tau, Z_\tau) \mathrm{d}W_\tau \mathrm{d}Z_\tau \cdots \mathrm{d}W_1 \mathrm{d}Z_1. \tag{8.10}$$

The density function $g(\cdot)$ can readily be made conditional on those observed Z_t that are exogenous or fixed and known. The joint density $g(\cdot)$ need not be the distribution that individuals use to account for the uncertainty they perceive about the future; $g(\cdot)$ describes the stochastic properties of the variables unobserved by the econometrician.

The sets Θ_{jP} are usually quite complicated to calculate. A popular simplifying assumption is to presume that individuals cannot save. In this case, $A_t = 0$ and $C_t = W_t + Y_t$ in Eqs. (8.5) and (8.7). With these assumptions we see that

$$V_t^P = U(P_t, P_t W_t + Y_t, Z_t) + \kappa E_t[\max V_{t+1}^P], \tag{8.11}$$

where the second term on the right-hand side of this expression does not depend at all on P_t. This formulation greatly simplifies computation of both the value functions and the sets Θ_{jP}. To simplify computation further, researchers also often assume that the variables W_t and Z_t are serially (and sometimes contemporaneously) independent.

8.1.3. Multiple values of hours

We can extend the above model beyond the simple decision to participate by admitting a limited set of hours choices. The approach shares many of the attributes of the computational-simplification procedure described in Section 6.7, with the complication that we must infer the value function appropriate for evaluating options. To illustrate this approach in a lifecycle context, suppose a worker may choose among full-time work, part-time work, and no work in each period, with each option implying a prescribed number of hours. This finite set of hours choices yields a relatively small set of discrete states, say J states in each period, over which the lifetime utility function must be maximized. Let P_{jt} designate a zero-one dummy variable equaling unity if an individual selects option j hours in period t, and let W_{jt} denote the earnings received from this option.

The value function now becomes

$$V_t \equiv \sum_{j=1}^{J} P_{jt} V_t^j \equiv \sum_{j=1}^{J} P_{jt} V_t^j(A_t, W_{jt}, Z_t), \tag{8.12}$$

where

$$V_t^j = \max_{C_t} \Bigg[U(P_{jt}, C_t, Z_t). \\ + \kappa E_t \left(\max_{P_{j(t+1)}} \sum_{j=1}^{J} P_{j(t+1)} V_{t+1}^j((1+r)(A_t - C_t + W_{jt} + Y_t), W_{j(t+1)}, Z_{t+1}) \right) \Bigg].$$

One can express V_t^j in a way similar to Eq. (8.5) which assists in computing value functions in many instances. The value function in the last period, τ, is

$$V\tau^j = \max_{C_\tau} U(P_{j\tau}, C_\tau, Z_\tau) \qquad \text{s.t.} \qquad C_\tau = A_\tau + W_{j\tau} + Y_\tau. \tag{8.13}$$

A backward recursive solution once again permits computation of each period's value functions and optimal choices.

Developing the likelihood function for the P_{jt}'s requires partitioning the sample space of W_{jt}'s and Z_t's corresponding to the particular decisions. Within period t, the decisions P_{jt} are mutually-exclusive and exhaustive. For notational convenience, suppose W_t now denotes a vector including all of the W_{jt}'s as elements. Define the sets

$$\Theta_{tj} = \{(W_t, Z_t) : V_t^j > V_t^k \text{ for } k = 1, ..., J, k \neq j\}. \tag{8.14}$$

When $(W_t, Z_t) \in \Theta_{tj}$, the individual chooses $P_{jt} = 1$. The set Θ_{tj} are functions of all decisions and variables observed in periods $t-1, t-2, ..., 1$.

The likelihood function for this more general case is given by Eq. (8.10), with the sets Θ_{tj} now replacing the sets Θ_{tP_t}. With this modification, $l_{P_1 P_2 ... P_T}$ represents the probability of observing the event $(P_{j1}, P_{j2}, ..., P_{jT})$.

Allowing for continuous choices in a lifecycle model involves insurmountable computational burden when participation is an issue, unless one relies on very strong behavioral and stochastic assumptions. In effect, this amounts to expanding the set J to a large number of values. Even in the simple case considering only participation, the above discussion shows that the entire lifecycle problem must be solved to characterize decisions in any period. The two-stage budgeting and Euler-condition approaches utilized in Section 4 are of little use in simplifying the estimation problem. Other behavioral features of lifecycle models diminish the usefulness of these approaches as well by invalidating the separability properties needed by them, even when participation is not a source of violation.

8.2. Learning by doing and human capital

Saving and the accumulation of assets is just one way that past labor supply choices can affect today's decisions. In learning-by-doing models, past work experience has a direct effect on the determination of market wages. A similar mechanism operates in human capital models. Past labor market decisions have an impact not just through the level of accumulated assets but also through the wage. These considerations significantly change the nature of the optimal labor supply decisions. For example, learning by doing introduces a trade-off between the increase in utility that can be achieved by reducing current work effort and the increase in future productivity that can be achieved from learning on the job. This implies that the current wage is no longer the appropriate measure of the return to working. An additional "dynamic rent" term must be included to account for increased future wages resulting from the accumulation of experience capital while working. Hence, the methods of Section 6, which are designed to deal with non-linearities in current wages arising from tax and transfer policies, are not directly useful here.

These dynamic generalizations of the standard model also imply that individuals who would have otherwise chosen to leave work may now choose to stay in employment. This property is also exhibited in search models that allow state dependence through asymmetry in layoff and job arrival rates. In this situation individuals may choose to remain in

employment so as to enhance the probability of being in employment when future returns to employment are high. For example, mothers of young children may choose to stay in employment simply to exploit the higher probability of subsequently being in work when children reach school age.

In the following we separate our discussion of these models into models with participation and those with continuous hours choices.

8.2.1. Learning by doing

8.2.1.1. Learning by doing with participation The learning-by-doing model posits that wages grow with experience. Individuals in these models do not decide whether or not to engage in human capital investment, the simple state of being in employment generates returns in its own right. The wage, W_t, is now determined as a function of experience capital, K_t. Experience capital in turn depends positively on past participation through a dynamic equation of the form

$$K_{t+1} = G(K_t, P_t). \tag{8.15}$$

Wages depend positively on K_t according to the function

$$W_t = W_t(K_t, \eta_t), \tag{8.16}$$

where η_t represents the unobservable component of wages as in Eq. (6.34). This implies that work not only brings immediate returns, but also increases future wages by adding to experience. For simplicity, we assume that the only uncertainty in the model enters through the wage error η_t.

The value functions in the period-t participation decision, corresponding to Eq. (8.4) have the form $V^P_{t+1}(A_{t+1}, W_{t+1}, K^P_{t+1}, \eta_{t+1}, X_{t+1})$, where X_t represents the elements of Z_{t+1} that remain after removing the K^P_{t+1} and η_{t+1} variables; now X_t incorporates all non-wage variables relevant for lifecycle decision making that are not controlled by the decision maker. This is done to explicitly acknowledge the dependence of K_{t+1} on P_t, and also to separate out the source of uncertainty η_{t+1}.

The solution to the individual's participation problem follows closely that outlined in Section 8.1. In period t, individuals choose participation to maximize utility as described by Eq. (8.4) but acknowledging the impact of P_t on K_{t+1} in V^P_{t+1}. Since the only uncertainty enters through η, the participation decision defines a "reservation value" for the wage error η^*_τ, which in turn defines the sets (8.8). This reservation value depends on the value of K_τ and thus, to solve the problem a solution must be found for each of the τ possible values of accumulated work experience. The definition of η^*_t for all periods and all possible value of K_t captures all of the economics of the problem. In each period, the individual realizes a wage shock and makes a work decision to maximize utility given accumulated experience. Accumulated experience impacts the decision both by increasing wages and by impacting the disutility of work. The impact of current work decisions on future utility is accounted for by the EV_{t+1} terms – working today changes the value of η^*

tomorrow and, thus, impacts the probability of future work and expected future utility. Given the value of η^* for all periods and all possible values of K_t, estimation is straightforward.

Following Eq. (8.12) this model can be extended to allow for additional discrete states, for example, part-time and full-time participation. Particular functional forms for $G(\cdot)$ may be also chosen to allow for interactions between K_t, participation and hours of work. We return to a discussion of specific parameterizations in the review of empirical studies in Section 8.4.

8.2.1.2. Learning by doing with continuous hours choices Often in a learning-by-doing model, the level of hours of work, rather than participation alone, determines wage growth. To introduce learning by doing in a model with continuous hours choices, we abstract from the participation decision and replace Eq. (8.15) with

$$K_{t+1} = G(H_t, K_t), \tag{8.17}$$

where H_t is hours of work in period t, and G is an increasing function of H_t. Choices over hours and consumption are made by maximizing equation

$$V(A_t, W_t, K_t, \eta_t, X_t) = \max_{C_t, L_t} [U(C_t, L_t, X_t) + \kappa E_t(V(A_{t+1}, W_{t+1}, K_{t+1}, \eta_{t+1} X_{t+1})]. \tag{8.18}$$

Notice that the value function in period t is made a function of the beginning of period t experience capital K_t as well as the financial capital A_t and the maximization takes place subject to the accumulation equations for experience capital and asset capital. The Euler equation for consumption continues to hold. However, the first order conditions for the allocation of time generalize to account for the role of experience capital. Assuming an interior solution for this continuous hours problem we have

$$U_L(C_t, L_t, X_t) = \lambda_t W_t + \kappa E_t\{\Gamma_{t+1}(\partial G/\partial H_t)\}, \tag{8.19}$$

$$\Gamma_t = \lambda_t(\partial W/\partial K_t)H_t + \kappa E_t\{\Gamma_{t+1}(\partial G/\partial K_t)\}, \tag{8.20}$$

where $\lambda_t = \partial V_t/\partial A_t = \partial U_t/\partial C_t$ and $\Gamma_t = \partial V_t/\partial K_t$.

The basic change from the standard hours of work model discussed in Section 4, is that the value of work is no longer simply the wage, but now includes the return to experience. This return depends on all future work decisions through the term Γ_{t+1} which measures the return to human capital. As such, standard hours of work equations of the sort we have been considering are inappropriate, in that they relate work to current wage which is no longer relevant on its own. All future wages and implied work decisions must also be included in determining the value of work.

8.2.2. Human capital

8.2.2.1. Human capital models with participation Consider an individual who, in each period, can now choose between participation in work, P_t, and participation in human

capital investment, P_t^*. The wage, W_t, is determined as a function of human capital, K_t, according to the function, $W_t = W_t(K_t, \eta_t)$ where K_t depends on past investment decisions. Suppose human capital accumulates according to the dynamic equation:

$$K_{t+1} = G(K_t, P_t^*). \tag{8.21}$$

The problem is, in principle, more complicated now because the individual must choose among three activities. However, this problem can be solved applying the multiple values of hours formulation outlined in Section 8.1.3. This is done simply by reinterpreting the discrete hours choices as options over the four states characterized by the four possible values combinations of P_t and P_t^*. There is the added complication that the wage is state dependent, but this is readily handled within Eqs. (8.12) and (8.13) given the definitions of the value functions V_t^j. Typically, the applications make the additional assumption of no savings to take attain computational simplifications. We discuss particular specifications in our review of the empirical applications at the end of this section.

8.2.2.2. Human capital models with continuous hours choices In the continuous hours-of-work problem individuals choose how much time in each period to spend in three activities: leisure L_t, hours of work H_t, and human capital investment S_t. Their choice problem is to choose L_t and C_t,

$$V(A_t, W_t, K_t, \eta_t, X_t) = \max_{C_t, L_t} [U(C_t, L_t, X_t) + \kappa E_t V(A_{t+1}, W_{t+1}, K_{t+1}, \eta_{t+1}, X_{t+1})] \tag{8.22}$$

subject to the human capital equations $K_{t+1} = G(K_t, S_t)$, the asset accumulation conditions and $L_t + H_t + S_t = T$. This results in two additional conditions:

$$U_L(C_t, L_t, X_t) = \kappa E_t\{\Gamma_{t+1}(\partial G/\partial S_t)\}, \tag{8.23}$$

$$\Gamma_t = \lambda_t H_t(\partial W_t/\partial K_t) + \kappa E_t\{\Gamma_{t+1}(\partial G/\partial K_t)\}. \tag{8.24}$$

Given these expressions, the marginal utility of leisure still equals λ times the wage rate, the marginal rate of substitution between consumption and leisure still equals W, and the Euler equation for consumption continues to hold.

Since time must be allocated among 3 activities, this problem becomes more complicated. Eq. (8.23) indicates that the return to training must also equal the return to leisure which equals the return to work. The return to schooling depends on the marginal value of a unit of human capital, Γ, and Eq. (8.24) gives a Euler equation for its time path. So, levels of both leisure and training must be selected to equate their marginal values with λ times the wage – these two choices together imply the number of hours worked. However, if hours of work can be measured separately from hours of training then the labor supply equation can be estimated directly since the standard marginal conditions for the choice of working hours remain valid.

8.3. Habit persistence

Habit persistence nullifies the intertemporal separability property for preferences through the dependence of current utility on past labor supply and consumption choices. In the framework introduced in Section 8.1, we can think of these past choices entering Z_t. For example, period-t utility may be written as $U_t(P_t, C_t, P_{t-1}, C_{t-1}, X_t)$. In this formulation we have divided Z_t into one set of elements controlled by the individual's previous behavior, namely P_{t-1} and C_{t-1}, and a second set designated as X_t that are not influenced by the decision maker. Typically this is set up as a household production model in which past non-market (leisure) time and past consumption influence today's utility. Consequently, one may wish to add further lags of participation and consumption. Our review of empirical applications considers such specifications.

The problem for the consumer is analogous to that described in Eq. (8.4), but now recognizing that Z_{t+1} is a function of current consumption and current participation. The backward recursion follows the same form as Eqs. (8.6) and (8.7). With the wage innovation η_t being the only source of uncertainty, the estimation is the same as in the learning-by-doing model.

This analogy with the learning-by-doing model also holds in the continuous hours choice framework without participation. In this case $U_t(C_t,L_t)$ is replaced by $U_t(C_t, L_t, C_{t-1}, L_{t-1}, X_t)$. Further lags may be included without changing the basic intuition underlying this model. The optimization problem in that case becomes

$$V(A_t, C_{t-1}, L_{t-1}, X_t) = \max_{C_t, L_t} \{U(C_t, L_t, C_{t-1}, L_{t-1}, X_t) + \kappa E_t V(A_{t+1}, C_t, L_t, X_{t+1})\}. \quad (8.25)$$

The first-order conditions for an interior solution for leisure becomes

$$U_L(C_t, L_t, C_{t-1}, L_{t-1}, X_t) + \kappa E_t\{U_L(C_{t+1}, L_{t+1}, C_t, L_t, X_{t+1})\}$$

$$= \kappa E_t\{U_C(C_{t+1}, L_{t+1}, C_t, L_t, X_{t+1})(1 + r)W_t\}. \quad (8.26)$$

A similar relation exists for consumption. As in the learning-by-doing model the value of work is no longer simply the wage. Now it includes the dynamic rent in terms of the impact on future marginal utility.

8.4. Review of empirical results

8.4.1. The basic intertemporal labor supply model

There are many applications of the basic intertemporal labor supply model. These are generally extensions of the Heckman and MaCurdy (1980) and MaCurdy (1981) studies.[60] For example, Browning et al. (1985) work directly with Frisch labor supply equations (see Section 4.4.3) and use a Psuedo cohort approach on the time series of repeated cross sections on consumption and family labor supply available in the British Family Expen-

[60] See also Altonji (1982, 1986).

diture Survey. They do not allow for non-participation. Blundell et al. (1993) incorporate corner solutions in their study of intertemporal hours of work decisions among married women in the UK. They work directly with the marginal conditions (4.6) and (4.20). The within-period consumption-leisure choices are modeled using an Almost Ideal form for preferences. The Euler equation is then used to identify a Box-Cox monotonic transformation of within-period utilities (as also adopted in MaCurdy (1983)). Their results point to intertemporal (Frisch) labor supply elasticities for married women in the 0.5–1 range depending on demographic characteristics – women with younger children having the bigger elasticities. As expected, estimated Marshallian elasticities are quite a bit smaller, in the 0.2–0.5 range. The intertemporal elasticity of substitution for consumption is approximately 0.6 which suggests a moderate degree of risk aversion.

8.4.2. *Learning-by-doing models*

Shaw (1989) estimates a learning-by-doing model in her study of the continuous hours choices of a similar sample of men from the PSID. She selects 526 men in the 18–64 year age range during the period 1967–1980. As in the Hotz et al. (1988) study, a Translog direct utility is chosen but this is specified in terms of *current* non-market time and consumption. There are no habit terms. However, in contrast to that earlier study, the stock of experience enters the wage equation. A quadratic specification is used for the capital accumulation function (8.17) to reflect the possibly concave nature of the lifecycle earnings profile. This is then used to define an estimable dynamic wage equation by assuming $W_t = \rho_t K_t$ where ρ_t is the rental rate of experience capital. This rental rate is assumed constant across individuals in any particular year. Shaw again finds strong evidence of non-separability – this time entering through the wage experience relationship rather than through the utility function. She finds a large positive effect which implies that a temporary 25% increase in hours of work increases wages by 12.8% starting from the initial mean values. The Shaw study is restricted to men and does not consider the problem of non-participation.

This is tackled in the Eckstein and Wolpin (1989) study which estimates a discrete model using a sample of 318 women from the NLS of mature women survey. They specify within-period utility to have the form (8.27). To simplify the problem, they assume that there is no saving or borrowing, so the within-period budget constraint reduces to

$$C_t = W_t P_t + Y_t. \tag{8.28}$$

Wages are assumed to be log linear in schooling, experience capital and the unobservable η, with $\eta_t \sim \text{iid}N(0, \sigma_\eta^2)$. Under these assumptions, the sample likelihood is given by[61]

[61] Obviously, the reservation wage cannot be bigger than the smallest wage observed for each individual of a particular type in the sample. Eckstein and Wolpin (1989) allow for measurement error in wages to avoid this restriction.

$$\prod_{i=1}^{N}\prod_{t=1}^{T_i}[\Phi(\eta_t^*/\sigma_\eta)]^{1-P_t}\left[\frac{1}{\sigma_\eta}\varphi(\eta_t/\sigma_\eta)\right]^{P_t}, \tag{8.29}$$

where η^* is derived from the structural utility maximization framework outlined in previous sections. Hence, the within-period problem is a standard Tobit formulation. Because the errors are serially uncorrelated, these within-period Tobit likelihood functions are simply multiplied together to yield the overall likelihood function.

The sample of women used in the Eckstein and Wolpin (1989) study were aged between 39 and 44 in 1967 and have at least four consecutive years of data on labor force participation beginning in 1966. The basic findings of the model are best summarized by the simulations the authors conduct, manipulating the value of each variable in the model and observing predicted work effort. First, they find that at any age, the probability of work increases with experience. Hence the positive impact of experience on wages overcomes the fact that the disutility of work increases with experience. Second, for any experience level, work effort decreases with age – as age advances there are less future gains available from increasing experience and, thus, the value of work declines. This explanation for declining work with age is missed by any static model. Third, work effort decreases with husband's earnings and increases with schooling. Finally, increasing the slope of the wage/experience profile substantially increases work effort over the lifetime. Again, this effect would be missed by any static labor supply model.

At this point, it is important to reiterate the extreme simplifying assumptions that have been made to make the problem manageable. First, a 0/1 work decision has been assumed. Second, individuals cannot save or borrow. These two assumptions together reduce the choice problem to a simple work, no-work decision, and limit the dynamic elements of the problem to the accumulation of human capital. Third, no unobserved heterogeneity is admitted in the utility function. The only error term in the model is the wage error, η, which is assumed serially uncorrelated and normally distributed. As we have seen, this reduces the dynamic problem to a series of standard Tobit problems and eliminates any concerns about initial conditions.

Altug and Miller (1990) combine certain aspects of both of these approaches in their study of labor supply and consumption. They use the Euler equation for consumption and the continuous hours information to recover some of the preference parameters. Utility is assumed explicitly additive in consumption and leisure but current-period utility is allowed to depend on past labor supply choices. Wages also have a multiplicative form in aggregate shocks, individual heterogeneity and a term capturing the effect of past labor supply choices. A log-differenced wage equation can, therefore, be estimated across individuals without adjustments for selection. To identify their model they are obliged to make certain additional assumptions on unobserved heterogeneity. First, they assume that, conditional on participation, there is no unobserved heterogeneity in hours of work. Second, they assume Pareto efficient allocations across all individuals in the economy. This latter assumption implies that the marginal utility of consumption is simply the

product of an individual and a time effect. A fixed cost of work parameter is introduced and recovered directly from the value function comparison. The forward looking terms in this comparison are simplified using the idea of Hotz and Miller (1993) which assumes sufficient stationarity to replace future value comparisons with current observed transition rates.

Estimation takes place using a sample of 2169 women from the PSID for 1973–1985. Consumption is restricted to food consumption. They find an important effect of past labor supply on wages. They also report important non-separabilities over time in utility. Current and past labor supplies are found to be substitutes.

8.4.3. Some extensions

This dynamic model has been extended in a number of papers to include endogenous fertility and marital decisions. For example, drawing on the earlier work of Heckman and Willis (1975) and Moffitt (1983), Hotz et al. (1988) develop a semi-reduced form representation of fertility and labor supply decision rules. Francesconni (1995) places this model in the Eckstein and Wolpin framework which is extended to allow endogenous fertility. Van Der Klaauw (1996) also presents an extension of this framework to allow for endogenous marital decisions, although he maintains the exogeneity of fertility.

Separability in the decision rule can also be relaxed through the introduction of asymmetric job layoff and arrival rates. This is the model presented in Blundell et al. (1997, 1998c) who developed earlier work on discouraged workers by Blundell et al. (1987) to allow for active search, layoffs and saving in a model of labor market transitions. Estimation is shown to be possible without recourse to the full dynamic programming solution using the information in the consumption Euler equation, labor market transition rates and the consumption policy function. However, strong restrictions are placed on the distribution of unobservable preference heterogeneity and on the distribution of wages.

8.4.4. Habit persistence models

The habit persistence model as discussed in Section 8.3 was investigated extensively in Hotz et al. (1988) although, as in the Shaw study, they do not consider non-participation. Their study further assumes that within-period utility over C_t and K_t in Eq. (8.26) is described by a Translog direct utility and they do not allow for learning by doing. Habits enter utility in the form

$$K_t = L_t + \alpha \Psi_t, \tag{8.30}$$

where Ψ_t is the habit stock of leisure

$$\Psi_t = (1 - \theta)\Psi_{t-1} + L_{t-1}. \tag{8.31}$$

The parameter α represents the substitution between current "leisure" and past leisure capital in the production of K_t. Notice that when the depreciation parameter, θ, in the definition of Ψ_t is unity then it is only last period's leisure (or labor supply) that matters for

today's marginal utility of income.[62] Allowing θ to be less than unity generalizes the first-order conditions slightly since now all future utilities depend on L_t through the stock term Ψ_t.

This specification results in two stochastic dynamic estimating equations which are estimated by generalized method of moments. Their application is to the hours and consumption choices of working men from the PSID panel for the period 1967–1978 (specifically 482 white household heads aged between 23 and 52). These two groups are subsequently split into a younger and older group. Although there is some evidence of misspecification in the consumption Euler equation, there is reasonably strong evidence of non-separable preferences and the parameters α and $1-\theta$ turn out to be precisely estimated at around 0.6 and 0.65, respectively, for the group of younger males who were aged 23–36 in 1967. For the sample of older men the α parameter is somewhat higher and the $1-\theta$ parameter slightly lower.

9. Closing comments

The aim of this chapter has been to critically review existing approaches to modeling labor supply and to identify important gaps in the literature that could be addressed in future research. We began with a look at the kind of policy reform proposals that labor supply models are now required to address and the set of labor market facts that labor supply models are designed to interpret. In the sections that followed, we developed a unifying framework and provided a brief assessment of each modeling approach, reviewing relevant empirical studies at the end of each section. In this concluding section, we ask: Have the recent advances in labor supply research, reviewed in this chapter, placed us in a better position to answer the policy reform questions raised in Section 2 and enabled us to provide a more reliable interpretation of the trends in participation and hours described in Section 3?

It is certainly true that this chapter has documented some significant advances in labor supply research since the original Handbook chapters on labor supply were written in the first half of the 1980s. Even relative to the important appraisal of the area by Heckman (1993), the marked changes in tax and welfare policies highlighted in Section 2 have forced labor supply research to increasingly acknowledge the importance of the extensive margin and discreteness in observed behavior. Likewise, the renewed focus on human capital in the policy debate has created the need for new generalizations in intertemporal models. We have also noted the innovations in our understanding of interactions between individuals within households concerning their labor supply decisions, brought about by the collective approach to family labor supply.

However, we have also identified some significant gaps in our knowledge which make it difficult to assert confidently that we are in a position to examine reliably many of the

[62] See, for example, the Johnson and Pencavel (1984) specification.

important current policy reform proposals or to assess accurately the main determinants of participation and hours-of-work changes. This, in turn, explains why labor supply remains an active and productive area for research. What are these gaps in our current knowledge? Perhaps the overriding difficulty remains with modeling participation. This is key in any analysis of welfare reform. Even in the simplest dynamic model without fixed costs, we have seen that the reservation wage depends on the whole future of wages and other unobservables. Estimation of behavioral parameters and simulation of policy reforms is, therefore, considerably complicated.

Some studies have attempted to restrict the margins for intertemporal decisions so as to focus on the discrete participation decision. Although these studies have provided important insights into modeling techniques and enhanced our understanding of behavior, it is difficult to believe that they provide sufficiently good approximations to actual behavior to give robust policy guidance. For example, we have argued that saving decisions should be modeled alongside labor supply decisions. Studies that model saving and labor supply allowing for discrete behavior are few and far between and no robust view of how these interactions work is currently available.

Some analysts have been content to measure the overall impact of past policy reforms on either participation or hours of work using a difference-in-differences or natural experiment approach. However, even where the stringent assumptions required for consistent estimation of interpretable parameters are satisfied, the estimated parameters do not provide sufficient information for extrapolation or simulation. We have argued that simulation of tax and welfare proposals cannot be completed without a structural model. Here a gap in the literature is revealed. Structural models allowing for discrete choices over labor-force and welfare participation that acknowledge dynamic decisionmaking are still not available in the empirical literature. A central part of this survey has been to assemble the building blocks necessary for such an analysis.

In a similar spirit, developments of the family labor supply model that allow for collective behavior must also be placed in an intertemporal context. Much evidence suggests that the strong pooling assumptions underlying the traditional family labor supply model are untenable, which is worrying for any analysis of the impact of welfare reform on family labor supply. However, only very recently have the simplest collective labor supply models been extended to allow for discrete choice and unobserved heterogeneity, both necessary ingredients of any empirical study. Moreover, allowing for the possibility of household production in these models requires more detailed data on time use.

Structural models that allow for the interactions between family members and the nonconvexities in the incentive structure facing individual workers typically place strong requirements on the individual's and the economist's knowledge of budget constraints and the distribution of unobservables. We have seen that it is often the desire for flexibility along these dimensions that motivates empirical studies that adopt the difference-in-differences approach. Even simple structural models often do not account for correlation between unobservable individual effects in labor supply and the wage and income vari

ables. Additionally, they do not allow for mismeasurement of the budget constraint or the wage and income variable themselves. Since structural models are required for many purposes for which labor supply analysis is undertaken, precisely how much these measurement issues matter for different datasets and different modeling approaches should remain an active area for research.

We have devoted much attention to the specification of labor supply models that account for non-convexities in the budget constraint, induced by high welfare withdrawal rates and fixed costs of work. This is no coincidence; the evaluation of the labor supply responses to welfare policy reforms remains the most significant recent contribution of standard labor supply models. We have already pointed out the need for further research that places this analysis in a dynamic setting. We have also noted the importance of research designed to assess the robustness of alternative approximations to the shape of the budget constraint and the packaging of hours choices into discrete bundles.

There remain a number of big issues that we have not touched on in this chapter but that are important for labor supply analysis. Many of these issues are discussed elsewhere in this Handbook. Among the most important is the modeling of the retirement decision. In a general sense, this is implicitly covered in our discussion of participation, but to properly understand the retirement decision requires careful treatment of the specific institutional structure of retirement programs and the way in which they interact with disability schemes and rules for earning after retirement (see the forthcoming volume by Gruber and Wise (1998) for a useful selection of country specific studies of the retirement behavior and the structure of social security systems). Another issue relates to the process of job search and job matching.

We should also acknowledge the potential importance of general equilibrium effects from tax and transfer programs. These make it even more difficult to think of groups of individuals wholly unaffected by reforms, as is required in the difference-in-differences approach, and imply different welfare calculations from those from models that assume gross wages and prices are unaffected by transfer and tax reforms.

Finally, we reiterate the main theme of our review: to formalize the assumptions that are required for interpretation of elasticities recovered from alternative modeling approaches and data sources. We hope that this has satisfied the twin goals of making clear precisely what is being estimated in any specific study and making it possible to compare estimates across studies.

Appendix A. Specifications of within-period preferences

This appendix briefly reviews some popular within-period (or contemporaneous or static) labor supply specifications.[63] Specification (4.30), used to illustrate our discussion in Section 4, corresponds to a within-period labor supply model of the form[64]

$$\ln H = \alpha \ln W + \theta Y + \rho. \tag{A.1}$$

Here we suppress the t subscript and allow the single quantity, ρ, to represent observed and unobserved heterogeneity. Specification (A.1) is one of a number of popular alternative three-parameter specifications that allow a single parameter for each of the wage, income effects and heterogeneity terms. Such models place strong restrictions on preferences and modern research on consumer behavior strives to relax these restrictions using more flexible representations.

One important restriction on preferences in within-period labor supply models is on the sign of the wage response. In theory there is no requirement for the wage response to be the same sign over all hours choices and, although it is required to be positive at the participation margin where the income effect is zero, it can become negative as hours increase. The precise shape of the hours – wage relationship is also likely to vary with income and demographic composition. A second restriction is on the income response, which determines the extent to which leisure is a normal good and whether it is a luxury or necessity. Most evidence from consumer behavior suggests that this varies widely across different goods and different types of consumers. Models that are linear in income (quasi-homothetic preferences) as in (4.30′), or that imply constant elasticities as in Eq. (4.30) are typically rejected.[65]

Restrictions on within-period preferences are usefully summarized by the specification of the indirect or direct utility function. The additivity between wage and income, implicit in (A.1), and the constancy of the wage elasticity for all hours choices, are reflected in the following additive exponential form of the indirect utility function:

$$v(W, Y) = \frac{W^{\alpha+1}}{\alpha + 1} - \frac{e^{-\theta Y}}{\theta e^{-\rho}}. \tag{A.2}$$

Many alternative three-parameter specifications of this kind are popular in empirical applications. The relationship among these specifications and the preference restrictions they imply are helpful in comparing studies. Here we list a number of them and provide a brief commentary.

Linear labor supply:

$$H = \alpha W + \theta Y + \rho \tag{A.3}$$

$$v(W, Y) = \exp(\theta W)\left(Y + \frac{\alpha}{\theta} W - \frac{\alpha}{\theta^2} + \frac{\rho}{\theta}\right). \tag{A.4}$$

[63] See Stern (1986) for a comprehensive review of these and more non-linear parametric static labor supply specifications and their implied indirect and direct utility functions.

[64] To provide a within-period interpretation of these preferences in a two-stage budgeting context, we would replace Y by the consumption-based measure, Y^C.

[65] The assumption of quasi-homothetic preferences provides a very poor approximation in empirical work on consumer behavior. More data-coherent specifications require terms not only in M_i but also in $M_i \ln(M_i)$ and even higher-order interactions.

Although popular, the linear model imposes the same sign on the wage response throughout and implies quasi-homothetic preferences.

Semi-log labor supply:

$$H = \alpha \ln W + \theta Y + \rho \tag{A.5}$$

$$v(W, Y; X) = \frac{\exp(\theta W)}{\theta}(\theta Y + \rho + \alpha \log W) - \frac{\alpha}{\theta}\int_{\theta W} \frac{\exp(\theta W)}{\theta W} \mathrm{d}(\theta W). \tag{A.6}$$

The semi-log model allows some non-linear curvature in wage effects so that the wage elasticity declines with hours but its sign is positive throughout and it is still linear in income. This formulation is attractive where non-participation is an issue and where there may be measurement error or endogeneity in wages and income. The log linearity in wage allows proportional taxes to enter linearly and is also a popular specification for reduced forms for gross hourly wages.

Semi-log labor supply (generalization 1):

$$H = \alpha \ln W + \theta Y^{*} + \rho \tag{A.7}$$

with $Y^{*} = WH + Y - \alpha W(1 - \exp(-H/\alpha))$.There is no easy form for the indirect utility with this generalization of the semi-log model but it is interesting for a number of reasons. First, it can be rewritten as a specification for the log marginal rate of substitution function which is linear in H and Y^{*}. Therefore, it produces a particularly simple form for the reservation wage. Second, it permits negative wage responses as hours increase. As H tends to zero, it approaches the standard semi-log model (A.6).

Semi-log labor supply (generalization 2):

$$H = \alpha \ln W + \theta Y/W + \rho \tag{A.8}$$

$$v(W, Y) = \frac{W^{\alpha+1}}{\alpha + 1}\left(\frac{Y}{W}(1 + \theta)^2 + \alpha \ln W + \rho - \frac{\alpha}{(1 + \theta)}\right). \tag{A.9}$$

This generalization has the attraction of allowing a change in sign for the wage elasticity as Y is reduced, which would typically correspond to an increase in labor supply. It also facilitates the introduction of higher-order terms in $\ln W$. However, this specification retains the assumption of linearity in Y and introduces an awkward non-linearity in W.

Stone–Geary (LES) labor supply:

The direct utility is probably the most familiar characterization:

$$u(H, C) = [\theta \ln(\gamma_H - H) + (1 - \theta)\ln(C - \gamma_C)] \tag{A.10}$$

with labor supply

$$WH = (1 - \theta)\gamma_H W - \theta Y + \theta \gamma_C. \tag{A.11}$$

The Stone–Geary specification, although popular in early work on household behavior, has been used less frequently in recent years. It can allow negative wage responses but it

corresponds to a direct utility that is explicitly additive in hours and consumption. That is, the log marginal rate of substitution is additive in consumption and hours. It is also quasi-homothetic. Notice, however, that it is equivalent to the second generalization of the semilog model (A.8) with $\ln W$ replaced by $1/W$.

CES labor supply:

This is a useful generalization of the LES labor supply and corresponds to choosing a direct utility of the form

$$u(H, C) = [\theta(\gamma_H - H)^{-\mu} + (1 - \theta)(C - \gamma_C)^{-\mu}]^{-1/\mu}. \tag{A.12}$$

It also implies an additive log marginal rate of substitution function and, therefore, explicit additivity between consumption and labor supply. However, it generalizes the substitution patterns between consumption and hours, and allows negative wage responses.

References

Aaberge, R., J. Dagsvik and S. Strom (1995), "Labour supply responses and welfare effects of tax reforms", Scandinavian Journal of Economics 97: 635–659.

Abbott, M. and O. Ashenfelter (1976), "Labor supply, commodity demand and the allocation of time", Review of Economic Studies 42: 389–411.

Ackum-Agell, S. and C. Meghir (1995), "Male labour supply in sweden: are incentive effects important?" Tax reform evaluation report no.12 (National Institute of Economic Research).

Agell, J., P. Englund and J. Sodersten (1996), "Tax reform of the century – the Swedish experiment", National Tax Journal, in press.

Altonji, J.G. (1982), "The intertemporal substitution model of labour market fluctuations: an empirical analysis", Review of Economic Studies 49: 783–824.

Altonji, J.G. (1986), "Intertemporal substitution in labor supply: evidence from micro data", Journal of Political Economy 94 (part II): S176–S215.

Altug, S. and R. Miller (1990), "Household choices in equilibrium", Econometrica 58: 543–570.

Amemiya, T. and T. MaCurdy (1986), "Instrumental variable estimation of an error components model", Econometrica 54: 869–881.

Angrist, J. (1991), "Grouped data estimation and testing in simple labor supply models", Journal of Econometrics 47: 243–265.

Apps, Patricia, F. and Ray Rees (1988), "Taxation and the household", Journal of Public Economics 35: 355–369.

Apps, Patricia F. and Ray Rees (1993), "Labor supply, household production and intra-family welfare distribution", Discussion paper (Australian National University).

Apps, Patricia F. and Ray Rees (1997), "Collective labor supply and household production", Journal of Political Economy 105: 178–190.

Arellano, M. and C. Meghir (1992), "Female labour supply and on-the-job search: an empirical model estimated using complementary data sets", Review of Economic Studies 59: 537–559.

Arrufat, J.L. and A. Zabalza (1986), "Female labour supply with taxation, random preferences, and optimization errors", Econometrica 54: 47–63.

Ashenfelter, O. (1983), "Determining participation in income-tested social programs", Journal of the American Statistical Society 78: 517–525.

Ashenfelter, O. and J. Ham (1979), "Education, unemployment and earnings", Journal of Political Economy 87 S99–S166.

Ashenfelter, O. and J.J. Heckman (1974), "The estimation of income and substitution effects in a model of family labor supply", Econometrica 42: 73–85.

Attanasio, O.P. and T. MaCurdy (1997), "Interactions in family labor supply and their implications for the impact of EITC", Mimeo. (Stanford University, Stanford, CA).

Balestra, P. and M. Nerlove (1966), "Pooling cross section and time series data in the estimation of a dynamic model : the demand for natural gas", Econometrica 34: 585–612.

Bingley, P. and I. Walker (1997), "The labor supply, unemployment and participation of lone mothers in in-work transfer programmes", Economic Journal 107: 1375–1390.

Bingley, P., G. Lanot, E. Symons and I. Walker (1995), "Child support reform and the labor supply of lone mothers in the UK", Journal of Human Resources 30: 256–279.

Blau, D. and P. Robins (1988), "Child-care costs and family labor supply", Review of Economics and Statistics 70: 374–381.

Blomquist, N.S. (1983), "The effect of income taxation on the labour supply of married men in Sweden", Journal of Public Economics 22: 169–197.

Blomquist, N.S. (1985), "Labour supply in a two-period model: the effect of a non-linear progressive income tax", Review of Economic Studies 52: 515–529.

Blomquist, S. (1996), "Estimation methods for male labour supply functions: how to take account of nonlinear taxes", Journal of Econometrics 70: 383–405.

Blomquist, S. and U. Hansson-Brusewitz (1990), "The effect of taxes on male and female labour supply in Sweden", Journal of Human Resources 25: 317–357.

Blomquist, S. and W. Newey (1997), "Nonparametric estimation of labor supply functions generated by piece wise linear budget constraints", Working paper no. 1997:24 (Department of Economics, Uppsala University, Sweden).

Blundell, R.W. (1986), "Econometric approaches to the specification of life-cycle labour supply and commodity demand behaviour", Econometric Reviews 6(1): 147–151.

Blundell, R.W. and P. Johnson (1998), "Pensions and labor market participation in the UK", American Economic Review (Papers and Proceedings) 88: 168–172.

Blundell, R.W. and C. Meghir (1986), "Selection criteria for a microeconometric model of labour supply", Journal of Applied Econometrics 1: 55–81.

Blundell, R.W. and C. Meghir (1987), "Bivariate alternatives to the tobit model", Journal of Econometrics 34: 179–200.

Blundell, R.W. and I. Walker (1982), "Modelling the joint determination of household labour supplies and commodity demands", Economic Journal 92: 58–74.

Blundell, R.W. and I. Walker (1986), "A life cycle consistent empirical model of labour supply using cross section data", Review of Economic Studies 53: 539–558.

Blundell, R.W., C. Meghir, E. Symons and I. Walker (1986), "A labour supply model for the simulation of tax and benefit reforms", in: R.W. Blundell and I. Walker, eds., Unemployment, search and labour supply (Cambridge University Press, Cambridge, UK).

Blundell, R.W., J. Ham and C. Meghir (1987), "Unemployment and female labour supply", Economic Journal 97: 44–64.

Blundell, R.W., C. Meghir, E. Symons and I. Walker (1988), "Labour supply specification and the evaluation of tax reforms", Journal of Public Economics 36: 23–52.

Blundell, R.W., V. Fry and C. Meghir (1990), "Preference restrictions in microeconometric models of life-cycle behaviour under uncertainty", in: J.P. Florens, ed., Microeconometrics: surveys and applications (Blackwell, Oxford, UK).

Blundell, R.W., A. Duncan and C. Meghir (1992), "Taxation and empirical labour supply models: lone parents in the UK", Economic Journal 102: 265–278.

Blundell, R., C. Meghir and P. Neves (1993), "Labour supply: an intertemporal substitution", Journal of Econometrics 59: 137–160.

Blundell, R.W., M. Browning and C. Meghir (1994), "Consumer demand and the life-cycle allocation of household expenditure", Review of Economic Studies 161: 57–80.
Blundell, R.W., T. Magnac and C. Meghir (1997), "Savings and labor market transitions", Journal of Business and Economic Statistics LXXIX: 527–539.
Blundell, R.W., P.-A. Chiappori, T. Magnac and C. Meghir (1998a), "Collective labor supply and participation", Working paper W98/20 (Institute for Fiscal Studies).
Blundell, R.W., A. Duncan and C. Meghir (1998b), "Estimating labour supply responses using tax policy reforms", Econometrica 66: 827–861.
Blundell, R.W., J. Ham and C. Meghir (1998c), "Unemployment, discouraged workers and female labour supply", Research in Economics 52: 103–131.
Boskin, Michael J. and Eytan Sheshinski (1983), "Optimal tax treatment of the family", Journal of Pubic Economics 20: 281–297.
Bourguignon, F. and T. Magnac (1990), "Labour supply and taxation in France", Journal of Human Resources 25: 358–389.
Bourguignon, F., M. Browning, P.-A. Chiappori and V. Lechene (1994), "Incomes and outcomes: a structural model and some evidence from French data", Journal of Political Economy 102: 1067–1096.
Borsh-Supan, A. and R. Schnabel (1998), "Social security and declining labor force particiaption in Germany", American Economic Review 88: 173–178.
Bover, O. (1989), "Estimating intertemporal labour supply elasticities using structural models", Economic Journal 99: 1026–1039.
Browning, M. and C. Meghir (1991), "Testing for separability between goods and leisure using conditional demand systems", Econometrica 59: 925–952.
Browning, M., A. Deaton and M. Irish (1985), "A profitable approach to labor supply and commodity demand over the life cycle", Econometrica 53: 503–543.
Burtless, G. and J. Hausman (1978), "The effect of taxes on labour supply", Journal of Political Economy 86: 1103–1130.
Card, D. (1994), "Intertemporal labour supply: an assessment", in: C. Sims, ed., Advances in econometrics (Cambridge University Press, Cambridge, UK).
Card, D. and P.K. Robins (1996), "Do financial incentives encourage welfare participants to work? Initial 18-month findings from the self-sufficiency project", (Social Research Corporation, Vancouver, BC, Canada).
Chiappori, Pierre-Andre (1988), "Rational household labor supply", Econometrica 56: 63–90.
Chiappori, Pierre-Andre (1992), "Collective labor supply", Journal of Public Economics 437–467.
Chiappori, Pierre-Andre (1997), "Introducing household production in collective models of labor supply", Journal of Political Economy 105: 191-209.
Cogan, J. (1980), "Labor supply with costs of labor market entry", in: James Smith, ed., Female labor supply: theory and estimation (Princeton University Press, Princeton, NJ) pp. 327–364.
Cogan, J.F. (1981), "Fixed costs and labor supply", Econometrica 49: 945–964.
Colombino, U. and D. Del Boca (1990), "The effect of taxes and labour supply in Italy", Journal of Human Resources 25: 390–414.
Dustmann, C. and A. Van Soest (1997), "Wage structures in the private and public sectors in West Germany" Fiscal Studies, in press.
Eckstein, Z. and K. Wolpin (1989), "Dynamic labour force participation of married women and endogenous work experience", Review of Economic Studies 56: 375–390.
Eissa, N. (1995a), "Taxation and labor supply of married women: the Tax Reform Act of 1986 as a natural experiment", Working paper no. 5023 (NBER, Cambridge, MA).
Eissa, N. (1995b), "Labor supply and the Economic Recovery Tax Act of 1981", in: M. Feldstien and J. Poterba, eds., Empirical foundations of household taxation, conference volume (NBER, Cambridge, MA).
Eissa, N. (1996), "Tax reforms and labor supply", in: James Poterba, ed., Tax policy and the economy, Vol. 10 (NBER, Cambridge, MA).

Eissa, N. and J. Liebman (1995), "Labor supply responses to the earned income tax credit", Working paper no. 5158 (National Bureau of Economic Research, Cambridge, MA).

Feldstein, M. (1995), "The effect of marginal tax rates on taxable income: a panel study of the 1986 Tax Reform Act", Journal of Political Economy 103: 551–572.

Flood, L.R. and T. MaCurdy (1992), "Work disincentive effects of taxes: an empirical analysis of Swedish men", Carnegie-Rochester Conference Series on Public Policy 37: 239–278.

Flood, L.R., A. Klevmarken et al. (1997), in: P. Olovsson, ed., Houshold market and nonmarket activities (HUS), Vols. 3–6 (Department of Economics, Uppsala University, Uppsala, Sweden).

Fortin, Bernard and Guy Lacroix (1997), "A test of neoclassical and collective models of household labor supply", Economic Journal 107: 933–955.

Fraker, T. and R. Moffitt (1988), "The effect of food stamps on labor supply: a bivariate selection model", Journal of Public Economics 35: 25–56.

Fraker, T., R. Moffitt and D. Wolf (1985), "Effective tax rates and guarantees in the AFDC program, 1967–1982", Journal of Human Resources 20: 251–263.

Francesconni, M. (1998), "A joint dynamic model of fertility and work of married women", Working paper of the ESRC Research Centre on Micro-Social Change, 98-2 (University of Essex, Colchester).

Friedberg, L. (1995), "The labor supply effects of the social security earnings test", Mimeo. (MIT, Boston, MA).

Gorman, W.M. (1959), "Separable utility and aggregation", Econometrica 21: 63–80.

Gorman, W.M. (1968), "The structure of utility functions", Revue of Economic Studies 32: 369–390.

Graversen, E.K. (1996), "Measuring labour supply responses to tax changes by use of exogenous tax reforms", Working paper no. 96-17 (Centre for Labour Market and Social Research, University of Aarhus, Aarhus, The Netherlands).

Gruber, J. and J. Wise (1998), "Social security and retirement: an international comparison", American Economic Review 88: 158–163.

Hall, R. (1973), "Wages, income and hours of work in the US labor force", in: G. Cain and H. Watts, eds., Income maintenance and labor supply (Chicago University Press, Chicago, IL).

Ham, J. (1986a), "Testing whether unemployment represents life-cycle labor supply behaviour", Review of Economic Studies LIII: 559–578.

Ham, J. (1986b), "On the interpretationof unemployment in empirical labour supply analysis", in: R.W. Blundell and I. Walker, eds., Unemployment, search and labour supply (Cambridge University Press, Cambridge, UK) pp. 121–142.

Hanoch, G. (1980), "A multivariate model of labor supply: methodology and estimation", in: James Smith, ed., Female labor supply: theory and estimation (Princeton University Press, Princeton, NJ).

Hausman, J. (1980), "The effect of wages, taxes and fixed costs on women's labor force participation", Journal of Public Economics 14: 161–194.

Hausman, J. (1981), "Labor supply. how taxes affect economic behavior", in: H. Aaron and J. Pechman, eds., Tax and the economy (Brookings Institution, Washington, DC).

Hausman, J. (1985a), "The econometrics of noninear budget sets", Econometrica 53: 1255–1282.

Hausman, J. (1985b), "Taxes and labor supply", in: A. Auerbach and M. Feldstein, eds., Handbook of public economics, Vol. 1 (North Holland, Amsterdam).

Hausman, J. and P. Ruud (1986), "Family labor supply with taxes", American Economic Review (Papers and Proceedings) 74: 242–248.

Hausman, J. and W. Taylor (1981), "Panel data and unobservable effects", Econometrica 49: 1377–1398.

Heckman, J.J. (1974a), "Shadow prices, market wages and labor supply", Econometrica 42: 679–694.

Heckman, J.J. (1974b), "Life-cycle consumption and labor supply: an explanation of the relationship between income and consumption over the life cycle", American Economic Review 64: 188–194.

Heckman, J.J. (1974c), "Effects of child-care programs on women's work effort", Journal of Political Economy 82(2): S136–S163.

Heckman, J.J. (1976), "Life-cycle model of earnings, learning and consumption", Journal of Political Economy 84: S11–S44.

Heckman, J.J. (1978), "Dummy endogenous variables in a simultaneous equation system", Econometrica 46: 931–959.

Heckman, J.J. (1979), "Sample selection bias as a specification error", Econometrica 47: 153–162.

Heckman, J.J. (1993), "What has been learned about labor supply in the past twenty years", American Economic Review (Papers and Proceedings) 83: 116–121.

Heckman, J.J. and T.E. MaCurdy (1980), "A life-cycle model of female labour supply", Review of Economic Studies 47: 47–74.

Heckman, J.J. and B. Singer (1984), "A method for minimizing the impact of distributional assumptions in econometric models for duration data", Econometrica 52: 271–320.

Heckman, J.J. and R.J. Willis (1977), "A beta-logistic model for the analysis of sequential labor force participation by married women", Journal of Political Economy 85: 27–58.

Hotz, V.J. and R.A. Miller (1993), "Conditional choice probabilities and the estimation of dynamic models", Review of Economic Studies 60: 497–530.

Hotz, V.J., F.E. Kydland and G.L. Sedlacek (1988), "Intertemporal substitution and labor supply", Econometrica 56: 335–360.

Hoynes, H.W. (1996), "Welfare transfers in two-parent families: labor supply and welfare participation under AFDC-UP", Econometrica 64(2): 295–332.

Ilmakunnas, S. and S. Pudney (1990), "A model of female labour supply in the presence of hours restriction", Journal of Public Economics 41: 183–210.

Johnson, T.R. and J.H. Pencavel (1984), "Dynamic hours of work functions for husbands, wives and single females", Econometrica 52: 363–389.

Kaiser, H., U. van Essen and P.B. Spahn (1992), "Income taxation and the supply of labour in West Germany", Jahrbucher fur Nationalokonomie und Statistik 209/1-2: 87–105.

Kapteyn, A. and I. Woittiez (1998), "Social interactions and habit formation in a model of female labour supply", Journal of Public Economics 70: 185–205.

Keane, M.P. and R. Moffitt (1995), "A structural model of multiple welfae program participation and labor supply", International Economic Review 39(3): 553–589.

Kell, M. and J. Wright (1989), "Benefits and the labour supply of women married to unemployed men", Economic Journal Conference Supplement: 1195–1265.

Killingsworth, M. and J. Heckman (1986), "Female labor supply: a survey", in: O. Ashenfelter and R. Layard, eds., Handbook of labor economics, Vol.1 (North-Holland, Amsterdam) pp. 103–204.

Kooreman, P. and A. Kapteyn (1986), "Estimation of rationed and unrationed household labour supply functions using flexible functional forms", Economic Journal 96: 308–322..

Kooreman, P. and A. Kapteyn (1990), "On the empirical implementation of some game theoretic models of household labour supply", Journal of Human Resources 25: 584–598.

Lundberg, S. (1988), "Labor supply of husbands and wives", Review of Economics and Statistics 70: 224–235.

MaCurdy, T.E. (1981), "An empirical model of labour supply in a life-cycle setting", Journal of Political Economy 89: 1059–1085.

MaCurdy, T.E. (1983), "A simple scheme for estimating an intertemporal model of labor supply and consumption in the presence of taxes and uncertainty", International Economic Review 24: 265–289.

MaCurdy, T.E. (1985), "Interpreting empirical models of labour supply in an intertemporal framework with uncertainty", in: J.J. Heckman and B. Singer, eds., Longitudinal analysis of labour market data (Cambridge University Press, Cambridge, UK).

MaCurdy, T.E. (1992), "Work disincentive effects of taxes: a re-examination of some evidence", American Economic Review 82: 243–249.

MaCurdy, T., D. Green and H. Paarsch (1990), "Assessing empirical approaches for analyzing taxes and labour supply", Journal of Human Resources 25: 415–490.

Manser, M. and M. Brown (1980), "Marriage and household decision making", International Economic Review 21: 31–44.

McElroy, M.B. (1981), "Empirical results from estimates of joint labor supply functions of husbands and wives", in: R.G. Ehrenberg, ed., Research in labor economics, Vol. 4 (JAI Press, Greenwich, CT) pp. 53–64.

McElroy, M. (1990), "The empirical content of nash-bargained household behavior", Journal of Human Resources 25: 559–583.

Moffitt, R. (1983), "An economic model of welfare stigma", American Economic Review 73: 1023–1035.

Moffitt, R. (1986), "The econometrics of piecewise-linear budget constraints: survey and exposition of the maximum likelihood method", Journal of Business and Economic Statistics 4: 317–327.

Moffitt, R. (1992a), "Estimating the value of an in-kind transfer: the case of food stamps", Econometrica 57: 385–409.

Moffitt, R. (1992b), "Incentive effects of the US welfare system: a review", Journal of Economic Literature 15: 1–161.

Moffitt, R. (1993), "Identification and estimation of dynamic models with a time series of repeated cross sections", Journal of Econometrics 59: 99–124.

Moffitt, R. (1993), "The effect of work and training programs on entry and exit from the welfare caseload", Discussion paper no. 1025-93 (Institute for Research on Poverty, University of Wisconsin, Madison, WI).

Moffitt, R. and B. Wolfe (1992), "The effect of the medicaid program on welfare participation and labor supply", Review of Economics and Statistics 74: 615–626.

Mroz, T.A. (1987), "The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions", Econometrica 55: 765–800.

Nakamura, A. and M. Nakamura (1981), "A comparison of the labour force behaviour of married women in the United States and Canada, with special attention to the impact of income taxes", Econometrica 49: 451–489.

Nerlove, M. (1971), "A note on error components models", Econometrica 39: 359–382.

Newey, W.K., J.L. Powell and J.R. Walker (1990), "Semiparametric estimation of selection models: some empirical results", American Economic Review 80(2): 324–328.

Pagan, A. (1986), "Two stage and related estimators and their applications", Review of Economic Studies 53: 517–538.

Pencavel, J. (1986), "Labor supply of men: a survey", in: O. Ashenfelter and R. Layard, eds., Handbook of labor economics, Vol. 1 (North-Holland, Amsterdam) pp. 3–102.

Ransom, M.R. (1987), "An empirical model of discrete and continuous choice in family labor supply", The Review of Economics and Statistics 59: 465–472.

Ribar, D. (1992), "Child care and the labor supply of married women", Journal of Human Resources 27: 134–165.

Robins, P. (1985), "A comparison of the labor supply findings from the four negative income tax experiments", Journal of Human Resources 20: 567–582.

Rosen, H.S. (1978), "The measurement of excess burden with explicit utility functions", Journal of Political Economy 86: s121–s135.

Shaw, K. (1989), "Life-cycle labour supply with human capital accumulation", International Economic Review 30(2): 431–457.

Smith, J.P. (1977), Female labor supply: theory and estimation (Princeton University Press, Princeton, NJ).

Smith, R.J. and R.W. Blundell (1986), "An exogeneity test for the simultaneous equation tobit model", Econometrica 54: 679–685.

Stern, N. (1986), "On the specification of labor supply functions", in: R.W. Blundell and I. Walker, eds., Unemployment, search and labour supply (Cambridge University Press, Cambridge, UK) pp. 121–142.

Triest, R. (1990), "The effect of income taxation on labor supply in the US", Journal of Human Resources 25: 491–516.

Van Soest, A., I. Woittiez and A. Kapteyn (1990), "Labour supply, income taxes and hours restrictions in the Netherlands", Journal of Human Resources 25: 517–558.

Van Soest, A. (1995), "Structural models of family labor supply: a discrete choice approach", Journal of Human Resources, 30: 63–88.

Wagner, Gert G., Richard V. Burkhuser and Friederike Behringer (1993), "The English language public use file or the German socio-economic panel", Journal of Human Resources 28(2 Spring): 429–433.

Wales, T.J. and A. Wodland (1976), "Estimation of household utility functions and labor supply response", International Economic Review 17: 69–80.

Zabalza, A. (1983), "The CES utility function, nonlinear budget constraints and labor supply", Economic Journal 93: 312–330.

Chapter 28

THE ECONOMIC ANALYSIS OF IMMIGRATION

GEORGE J. BORJAS*

Harvard University

Contents

Abstract 1698
JEL codes 1698
1 Introduction 1698
2 Immigration and the host country's economy 1700
2.1 A model with homogeneous labor 1700
2.2 Heterogeneous labor and perfectly elastic capital 1703
2.3 Heterogeneous labor and inelastic capital 1705
2.4 Simulating the impact of immigration 1707
3 The skills of immigrants: theory 1709
3.1 The migration decision 1710
3.2 The self-selection of immigrants 1711
3.3 Selection in observed characteristics 1716
4 The skills of immigrants: empirics 1717
4.1 The identification problem 1718
4.2 Economic assimilation 1721
4.3 Empirical evidence for the United States 1722
4.4 Convergence and conditional convergence 1728
5 Immigration and the wage structure 1733
5.1 Spatial correlations 1734
5.2 A model of wage determination and internal migration 1740
5.3 A model with a permanent supply shock 1746
5.4 Immigration and native internal migration 1748
5.5 The factor proportions approach 1753
6 Conclusion 1755
References 1757

* Pforzheimer Professor of Public Policy, John F. Kennedy School of Government, Harvard University; and Research Associate, National Bureau of Economic Research. I am grateful to the National Science Foundation for research support.

Handbook of Labor Economics, Volume 3, Edited by O. Ashenfelter and D. Card

Abstract

The study of labor flows across labor markets is a central ingredient in any discussion of labor market equilibrium. These labor flows help markets reach a more efficient allocation of resources. This paper surveys the economic analysis of immigration. It investigates the determinants of the immigration decision by workers in source countries and the impact of that decision on the host country's labor market. The survey stresses the ideas and models that economists use to analyze immigration, and delineates the implications of these models for empirical research and for our understanding of the labor market effects of immigration.

JEL codes: J1; J3; J6

1. Introduction

Why do some people move? And what happens when they do? The study of labor flows across labor markets – whether within or across countries – is a central ingredient in any discussion of labor market equilibrium. These labor flows help markets reach a more efficient allocation of resources. As a result, the questions posed above have been at the core of labor economics research for many decades.

At the end of the 20th century, about 140 million persons – or roughly 2% of the world's population – reside in a country where they were not born.[1] Nearly 6% of the population in Austria, 17% in Canada, 11% in France, 17% in Switzerland, and 9% in the United States is foreign-born.[2] These sizable labor flows have altered economic opportunities for native workers in the host countries, and they have generated a great deal of debate over the economic impact of immigration and over the types of immigration policies that host countries should pursue.

This chapter surveys the economic analysis of immigration.[3] In particular, the study investigates the determinants of the immigration decision by workers in source countries and the impact of that decision on the labor market in the host country. There already exist a number of surveys that stress the implications of the empirical findings in the immigration literature, particularly in the US context (Borjas, 1994; Friedberg and Hunt, 1995; LaLonde and Topel, 1996). This survey also reviews the empirical evidence, but it differs by stressing the ideas and models that economists use to analyze immigration, and by delineating the implications of these models for empirical research and for our understanding of the labor market effects of immigration. A key lesson of economic theory is that the labor market impact of immigration hinges crucially on how the skills of immi-

[1] Martin (1998).

[2] United Nations (1989, p. 61).

[3] Although the discussion focuses on the economic analysis of international migration, many of the models and concepts can also be used to analyze migration behavior within a country. Greenwood (1975) surveys the extensive literature on internal migration decisions.

grants compare to those of natives in the host country. And, in fact, much of the research effort in the immigration literature has been devoted to: (a) understanding the factors that determine the relative skills of the immigrant flow; (b) measuring the relative skills of immigrants in the host country; and (c) evaluating how relative skill differentials affect economic outcomes.

Because the survey focuses on the impact of immigration on the host country's labor market, the analysis ignores a number of important and equally interesting issues – both in terms of their theoretical implications and of their empirical significance. Immigration, after all, affects economic opportunities not only in the host country, but in the source country as well. Few studies, however, investigate what happens to economic opportunities in a source country when a selected subsample of its population moves elsewhere. Immigration also has economic effects on the host country that extend far beyond the labor market. An important part of the modern debate over immigration policy, for instance, concerns the impact of immigrants on expenditures in the programs that make up the welfare state. Finally, the survey focuses on the economic impact of immigrants, and ignores the long-run impact of the children and grandchildren of immigrants on the host country.[4]

The survey is structured as follows. Section 2 examines how immigration affects labor market opportunities in the host country. Economic theory implies that immigrants will generally increase the national income that accrues to the native population in the host country, and that these gains are larger the greater the differences in productive endowments between immigrants and natives. Section 3 analyzes the factors that determine the skills of immigrants. The discussion summarizes the implications of the income-maximization hypothesis for the skill composition of the self-selected immigrant flow. Section 4 discusses the identification problems encountered by studies that attempt to estimate how the skills of immigrants compare to those of natives – both at the time of entry and over time as immigrants adapt to the host country's labor market. The discussion also examines the concept of economic assimilation and investigates the nature of the correlation between an immigrant's "pre-existing" skills and the skills that the immigrant acquires in the host country. Section 5 surveys the vast literature that attempts to measure the impact of immigration on the wage structure in the host country. For the most part this literature estimates "spatial correlations" – correlations between economic outcomes in an area (such as a metropolitan area or a state in the United States) and the immigrant supply shock in that area. The section presents a simple economic model to illustrate that these spatial correlations typically do not estimate any parameter of interest, and suggests how these spatial correlations can be adjusted to estimate the "true" wage effects of immigration as long as estimates of native responses to immigration are available. Finally,

[4] There is increasing interest in analyzing how the skill composition of the immigrant flow affects the skill distribution of the children and grandchildren of immigrants. Borjas (1992) finds that skill differentials across the national origin groups in the immigrant generation tend to persist into the second and third generations, and attributes part of this persistence to "ethnic externalities."

Section 6 offers some concluding remarks and discusses some research areas that require further exploration.

2. Immigration and the host country's economy

This section uses a simple economic framework to describe how immigration affects the labor market in the host country, and to calculate the gains and losses that accrue to different groups in the population.[5] The analysis shows that natives in the host country benefit from immigration as long as immigrants and natives differ in their productive endowments; that the benefits are larger the greater the differences in endowments; and that the benefits are not evenly distributed over the native population – natives who have productive endowments that complement those of immigrants gain, while natives who have endowments that compete with those of immigrants lose.

2.1. A model with homogeneous labor

Suppose the production technology in the host country can be summarized by a twice-differentiable and continuous linear homogeneous aggregate production function with two inputs, capital (K) and labor (L), so that output $Q = f(K, L)$. The work force contains N native and M immigrant workers, and all workers are perfect substitutes in production ($L = N + M$). Natives own the entire capital stock in the host country and, initially, the supply of capital is perfectly inelastic. Finally, the supplies of both natives and immigrants are also perfectly inelastic.[6]

In a competitive equilibrium, each factor price equals the respective value of marginal product. Let the price of the output be the numeraire. The rental rate of capital in the pre-immigration equilibrium is $r_0 = f_K(K, N)$ and the price of labor is $w_0 = f_L(K, N)$. Because the aggregate production function exhibits constant returns, the entire output is distributed to the owners of capital and to workers. In the pre-immigration regime, the national income accruing to natives, Q_N, is given by

$$Q_N = r_0 K + w_0 L. \tag{1}$$

Fig. 1 illustrates this initial equilibrium. Because the supply of capital is fixed, the area under the marginal product of labor curve (f_L) gives the economy's total output. The national income accruing to natives Q_N is given by the trapezoid ABN0.

The entry of M immigrants shifts the supply curve and lowers the market wage to w_1.

[5] Borjas (1995b) and Johnson (1997) present more extensive discussions of this framework. Benhabib (1996) gives a political economy extension that examines how natives form voting coalitions to maximize the gains from immigration.

[6] The calculation of the gains from immigration would be more cumbersome if native labor supply was not inelastic because the analysis would have to value the change in utility experienced by native workers as they move between the market and non-market sectors.

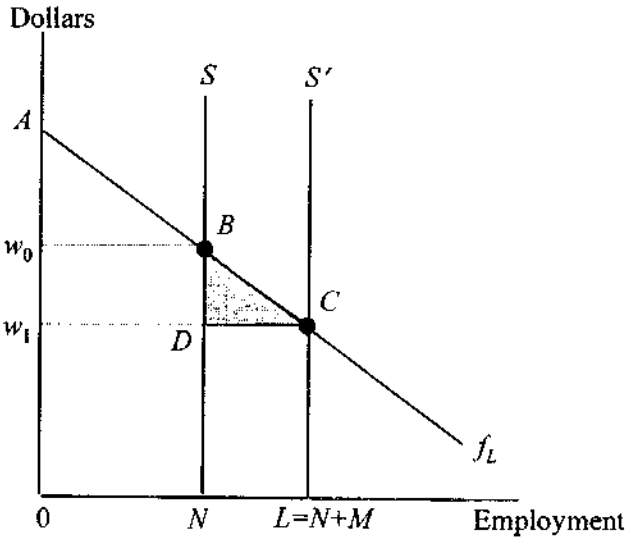

Fig. 1. The immigration surplus in a model with homogeneous labor and fixed capital.

The area in the trapezoid ACL0 now gives national income. Part of the increase in national income is distributed directly to immigrants (who get w_1M in labor earnings). The area in the triangle BCD gives the increase in national income that accrues to natives, or the "immigration surplus."

The area of BCD is given by $(1/2)\times(w_0 - w_1)\times M$. The immigration surplus, as a fraction of national income, equals[7]

$$\frac{\Delta Q_N}{Q} = -\frac{1}{2}\alpha_L \varepsilon_{LL} m^2, \tag{2}$$

where α_L is labor's share of national income ($\alpha_L = wL/Q$); ε_{LL} is the elasticity of factor price for labor ($\varepsilon_{LL} = \mathrm{d}\log w/\mathrm{d}\log L$, holding marginal cost constant); and m is the fraction of the work force that is foreign born ($m = M/L$).

Eq. (2) can be used to make "back-of-the-envelope" calculations of how much a host country gains from immigration. In the United States, the share of labor income is about 70%, and the fraction of immigrants in the work force is slightly less than 10%. Hamermesh's (1993, pp. 26–29) survey of the empirical evidence on labor demand suggests that the elasticity of factor price for labor may be around −0.3. The US immigration surplus, therefore, is on the order of 0.1% of GDP.

Eq. (2) shows that the immigration surplus is proportional to ε_{LL}. The net gains from immigration to the host country, therefore, are intimately linked to the adverse impact that

[7] The derivation in (2) uses the approximation that $(w_0 - w_1) \approx -(\partial w/\partial L)\times M$.

immigration has on the wage of competing native workers. If the increase in labor supply greatly reduces the wage, natives *as a whole* gain substantially from immigration. If the native wage does not respond to the admission of immigrants, the immigration surplus is zero.[8]

Immigration redistributes income from labor to capital. In terms of Fig. 1, native workers lose the area in the rectangle $w_0\mathrm{BD}w_1$, and this quantity plus the immigration surplus accrues to capitalists. Expressed as fractions of GDP, the net changes in the incomes of native workers and capitalists are approximately given by[9]

$$\left.\frac{\text{Change in native labor earnings}}{Q}\right|_{dK=0} = \alpha_L \varepsilon_{LL} m(1-m), \tag{3}$$

$$\left.\frac{\text{Change in income of capitalists}}{Q}\right|_{dK=0} = -\alpha_L \varepsilon_{LL} m\left(1-\frac{m}{2}\right). \tag{4}$$

Consider again the calculation for the United States. If the elasticity of factor price is -0.3, native-born workers lose about 1.9% of GDP, while native-owned capital gains about 2.0% of GDP. The small immigration surplus can disguise a sizable income redistribution from workers to the users of immigrant labor.

The derivation of the immigration surplus in Eq. (2) assumed that the host country's capital stock is fixed. However, immigrants may themselves add to the capital stock of the host country, and the rise in the return to capital will encourage capital flows into the country until the rental rate is again equalized across markets.[10]

As an alternative polar assumption, suppose that the supply of capital is perfectly elastic at the world price ($dr = 0$). Differentiating the marginal productivity condition $r = f_K(K, L)$ implies that the immigration-induced change in the capital stock is

$$\left.\frac{dK}{dM}\right|_{dr=0} = -\frac{f_{KL}}{f_{KK}} > 0. \tag{5}$$

The derivative in (5) is positive because $f_{KL} > 0$ when the production function is linear homogeneous. For convenience, assume that the additional capital stock defined by (5) either originates abroad and is owned by foreigners, or is owned by the immigrants themselves.

The elasticity of complementarity for any input pair i and j is $c_{ij} = f_{ij}f/f_i f_j$.[11] The

[8] The gains from immigration and the adverse impact on the native wage are directly linked unless all immigrants have skills that complement those of native workers.

[9] Eq. (3) uses the approximation that $(w_0 - w_1)N = -(\partial w/\partial L) \times M \times N$. The gains accruing to capitalists are calculated by adding the absolute value of this expression to the immigration surplus.

[10] However, Feldstein and Horioka (1980) find evidence that capital is somewhat immobile across countries.

[11] The elasticity of complementarity is the dual of the elasticity of substitution. Hamermesh (1993, Chapter 2) presents a detailed discussion of the properties of the elasticity of complementarity.

elasticity of factor price is proportional to the elasticity of complementarity, or $\varepsilon_{ij} = \alpha_j c_{ij}$, where α_j gives the share of income accruing to j. The immigration-induced wage change is given by

$$\left.\frac{\mathrm{d}\log w}{\mathrm{d}\log M}\right|_{\mathrm{d}r=0} = \left(\varepsilon_{LK}\left.\frac{\mathrm{d}\log K}{\mathrm{d}\log M}\right|_{\mathrm{d}r=0} + \varepsilon_{LL}m\right) = \frac{\alpha_L}{c_{KK}}(c_{KK}c_{LL} - c_{LK}^2)m. \tag{6}$$

The linear homogeneity of the production function implies that $c_{KK}c_{LL} - c_{LK}^2 = 0$, so that the host country's wage is independent of immigration. Hence the immigration surplus when the supply curve of capital is perfectly elastic is

$$\left.\frac{\Delta Q_N}{Q}\right|_{\mathrm{d}r=0} = 0. \tag{7}$$

The immigration-induced capital flow reestablishes the pre-immigration capital/labor ratio in the host country. Immigration does not alter the price of labor or the returns to capital, and natives neither gain nor lose from immigration.

2.2. Heterogeneous labor and perfectly elastic capital

Suppose there are two types of workers in the host country's labor market, skilled (L_S) and unskilled (L_U). The linear homogeneous aggregate production function is given by

$$Q = f(K, L_S, L_U) = f[K, bN + \beta M, (1 - b)N + (1 - \beta)M], \tag{8}$$

where b and β denote the fraction of skilled workers among natives and immigrants, respectively.[12] The production function is continuous and twice differentiable, with $f_i > 0$ and $f_{ii} < 0$ ($i = K, L_S, L_U$). The price of each factor of production, r for capital and w_i ($i = \mathrm{S}, \mathrm{U}$) for labor, is determined by the respective marginal productivity condition. As we saw earlier, the economic impact of immigration depends crucially on what happens to the capital stock when immigrants enter the country. Let's initially consider the case where the supply of capital is perfectly elastic, so that $\mathrm{d}r = 0$. Let p_S and p_U be the shares of the work force that are skilled and unskilled, respectively. The condition that $r = f_K(K, L_S, L_U)$ is constant implies that the immigration-induced adjustment in the capital stock equals

$$\left.\frac{\mathrm{d}K}{\mathrm{d}M}\right|_{\mathrm{d}r=0} = -\frac{[f_{KS}\beta + f_{KU}(1 - \beta)]}{f_{KK}}. \tag{9}$$

We can determine the impact of immigration on the wages of skilled and unskilled workers by differentiating the respective marginal productivity conditions, and by imposing the restriction in Eq. (9). The wage effects of immigration are:[13]

[12] A more general model would allow the host country to produce and consume more than one output. This generalization introduces additional sources of potential complementarity between immigrants and natives. The model, however, is much more complex. Trefler (1997) presents a discussion of these types of models in an open economy framework.

$$\left.\frac{d\log w_S}{d\log M}\right|_{dr=0} = \frac{\alpha_S}{c_{KK}}[c_{SS}c_{KK} - c_{SK}^2]\frac{(\beta - b)}{p_S p_U}(1 - m)m, \tag{10}$$

$$\left.\frac{d\log w_U}{d\log M}\right|_{dr=0} = \frac{-\alpha_U}{c_{KK}}[c_{UU}c_{KK} - c_{UK}^2]\frac{(\beta - b)}{p_S p_U}(1 - m)m, \tag{11}$$

where α_i is the share of national income accruing to factor i.

One can always write a linear homogeneous production function with inputs (X_1,X_2,X_3) as $Q = X_3 g(X_1/X_3, X_2/X_3)$. Suppose that the function g is strictly concave, so that the isoquants between any pair of inputs have the conventional convex shape. This assumption implies that $c_{11}c_{22} - c_{12}^2 > 0$. Eqs. (10) and (11) then indicate that the impact of immigration on the wage structure depends entirely on how the skill distribution of immigrants compares to that of natives. If the two skill distributions are equal ($\beta = b$), immigration has no impact on the wage structure of the host country. If immigrants are relatively unskilled ($\beta < b$), the unskilled wage declines and the skilled wage rises. If immigrants are relatively skilled ($\beta > b$), the skilled wage declines and the unskilled wage rises. In short, the impact of immigration on the wage structure depends on the *relative* skills of immigrants, not on their absolute skills.

The immigration surplus in this model is defined by

$$\left.\Delta Q_N\right|_{dr=0} = \left(bN\frac{\partial w_S}{\partial M} + (1 - b)N\frac{\partial w_U}{\partial M}\right)M. \tag{12}$$

It is well known that when the derivatives in (12) are evaluated at the initial equilibrium, where $L_S = bN$ and $L_U = (1 - b)N$, the infinitesimal increase in national income accruing to natives is zero.[14] To calculate finite changes, evaluate the immigration surplus using an "average" rate for $\partial w_S/\partial M$ and $\partial w_U/\partial M$, where the averages are defined by

$$\frac{1}{2}\left(\left.\frac{\partial w_S}{\partial M}\right|_{L_S = bN} + \left.\frac{\partial w_S}{\partial M}\right|_{L_S = bN + \beta M}\right),$$

and by

$$\frac{1}{2}\left(\left.\frac{\partial w_U}{\partial M}\right|_{L_U = (1-b)N} + \left.\frac{\partial w_S}{\partial M}\right|_{L_U = (1-b)N + (1-\beta)M}\right),$$

respectively.[15] By using Eqs. (10) and (11), it can be shown that the immigration surplus as a fraction of national income is given by[16]

[13] The derivation of Eqs. (10) and (11) is somewhat tedious and requires using the identities $(\varepsilon_{SS}\varepsilon_{KK} - \varepsilon_{SK}\varepsilon_{KS}) \equiv -(\varepsilon_{SU}\varepsilon_{KK} - \varepsilon_{SK}\varepsilon_{KU})$ and $(\varepsilon_{UU}\varepsilon_{KK} - \varepsilon_{UK}\varepsilon_{KU}) \equiv -(\varepsilon_{US}\varepsilon_{KK} - \varepsilon_{UK}\varepsilon_{KS})$. These identities follow from the fact that a weighted average of factor price elasticities equals zero.

[14] Bhagwati and Srinivasan (1983, p. 294).

[15] This approximation implies that the finite change in the immigration surplus is half the gain obtained when Eq. (12) is evaluated at the post-immigration level of labor supply.

$$\left.\frac{\Delta Q_N}{Q}\right|_{dr=0} = \frac{-\alpha_S^2}{2c_{KK}}[c_{SS}c_{KK} - c_{SK}^2]\frac{(\beta - b)^2}{p_S^2 p_U^2}(1-m)^2 m^2. \tag{13}$$

The immigration surplus is zero if $\beta = b$, and positive if $\beta \neq b$. If immigrants had the same skill distribution as natives, the immigration-induced change in the capital stock implies that the wages of skilled and unskilled workers are unaffected by immigration. The gains arise only if immigrants differ from natives.

Let β^* be the value of β that maximizes the immigration surplus in the host country. By partially differentiating Eq. (13) with respect to β, we obtain[17]

$$\begin{aligned} &\beta^* = 1, \qquad \text{if } b < 0.5, \\ &\beta^* = 0 \text{ or } \beta^* = 1, \qquad \text{if } b = 0.5, \\ &\beta^* = 0, \qquad \text{if } b > 0.5. \end{aligned} \tag{14}$$

Suppose that $b = 0.5$. There is no immigration surplus if half of the immigrant flow is also composed of skilled workers. The immigration surplus is maximized when the immigrant flow is either exclusively skilled or exclusively unskilled. Either policy choice generates an immigrant flow that is very different from the native work force.

Economic incentives for moving to a particular tail of the skill distribution arise when the native work force is relatively skilled or unskilled. Suppose the native work force is relatively unskilled ($b < 0.5$). Admitting skilled immigrants, who most complement native workers, maximizes the immigration surplus. If the native work force is relatively skilled, the host country should admit unskilled immigrants to maximize the gains.

2.3. Heterogeneous labor and inelastic capital

The results in (14) are very sensitive to the assumption that the supply curve of capital is perfectly elastic. Suppose instead that the capital stock is perfectly inelastic and is owned by natives. By differentiating the marginal productivity conditions, it can be shown that the changes in the various factor prices are given by

$$\left.\frac{d\log r}{d\log M}\right|_{dK=0} = \varepsilon_{KS}\frac{(\beta - b)}{p_S p_U}(1-m)m - \varepsilon_{KK}\frac{1-\beta}{p_U}m, \tag{15}$$

[16] The derivation of Eq. (13) uses the fact that $\alpha_S^2(c_{KK}c_{SS} - c_{SK}^2) = \alpha_U^2(c_{KK}c_{UU} - c_{UK}^2)$. This restriction follows from the identities defined in note 11.

[17] The differentiation assumes that the immigrant supply shock is "small" and does not affect the values of p_S and p_U.

$$\left.\frac{d\log w_S}{d\log M}\right|_{dK=0} = \varepsilon_{SS}\frac{(\beta - b)}{p_S p_U}(1 - m)m - \varepsilon_{SK}\frac{1 - \beta}{p_U}m, \tag{16}$$

$$\left.\frac{d\log w_U}{d\log M}\right|_{dK=0} = -\varepsilon_{UU}\frac{(\beta - b)}{p_S p_U}(1 - m)m - \varepsilon_{UK}\frac{1 - \beta}{p_U}m. \tag{17}$$

Immigration alters the distribution of income even when immigrants have the same skill distribution as natives. Suppose, in fact, that $\beta = b$. Eq. (15) then shows that immigration increases the rental rate of capital (ε_{KK} is negative). Moreover, immigration reduces the *total* earnings of native workers:

$$\text{Change in labor earnings}\Big|_{dK=0} = bN\frac{\partial w_S}{\partial M}M + (1 - b)N\frac{\partial w_U}{\partial M}M$$

$$= -Q_N[\alpha_S\varepsilon_{SK} + \alpha_U\varepsilon_{UK}]\frac{(1 - b)^2}{p_U^2}(1 - m)m < 0. \tag{18}$$

The sign of (18) follows from the fact that a weighted average of factor price elasticities equals zero ($\alpha_K\varepsilon_{KK} + \alpha_S\varepsilon_{SK} + \alpha_U\varepsilon_{UK} = 0$). Even though immigrants have the same skill distribution as natives, immigration reduces the capital/labor ratio and workers, as a group, lose.

The immigration surplus equals

$$\Delta Q_N\Big|_{dK=0} = \left(K\frac{\partial r}{\partial M} + bN\frac{\partial w_S}{\partial M} + (1 - b)N\frac{\partial w_U}{\partial M}\right)M. \tag{19}$$

By using the wage effects defined in Eqs. (15)–(17) and evaluating the various derivatives in (19) at the "average" point, we obtain

$$\left.\frac{\Delta Q_N}{Q}\right|_{dK=0} = -\frac{\alpha_S^2 c_{SS}\beta^2 m^2}{2p_S^2} - \frac{\alpha_U^2 c_{UU}(1 - \beta)^2 m^2}{2p_U^2} - \frac{\alpha_S\alpha_U c_{SU}\beta(1 - \beta)m^2}{p_S p_U}. \tag{20}$$

The quadratic form in (20) is positive.[18] Natives gain from immigration, therefore, even if the skill distribution of immigrants is the same as that of natives.

To illustrate the relationship between the immigration surplus and the skill distribution of immigrants, let V be the immigration surplus defined in (20) and consider the special case where $p_S = p_U = 0.5$. The first and second derivatives of the immigration surplus are proportional to

$$\frac{\partial V}{\partial \beta} \propto -\alpha_S^2 c_{SS}\beta + \alpha_U^2 c_{UU}(1 - \beta) - \alpha_S\alpha_U c_{SU}(1 - 2\beta), \tag{21}$$

[18] Eq. (20) is a quadratic form in the negative-definite matrix $\begin{bmatrix} c_{SS} & c_{SU} \\ c_{US} & c_{UU} \end{bmatrix}$.

$$\frac{\partial^2 V}{\partial \beta^2} \propto -\alpha_S^2 c_{SS} - \alpha_U^2 c_{UU} + 2\alpha_S \alpha_U c_{SU}. \tag{22}$$

Suppose now that $c_{SS} < c_{UU}$ (which implies that $\varepsilon_{SS} < \varepsilon_{UU}$ and the demand for skilled labor is less elastic than the demand for unskilled labor). This assumption tends to be supported by the empirical evidence (Hamermesh, 1993, Chapter 3). The first derivative is then positive at $\beta = 1$, and the second derivative is positive everywhere, so that (20) is convex.[19]

Evaluating the immigration surplus in Eq. (20) at $\beta = 0$ or $\beta = 1$, and using the convexity restrictions in (21) and (22), implies that the immigration surplus is maximized when the immigrant flow is exclusively skilled. The assumption that the wages of skilled workers are more responsive to a supply shift than the wages of unskilled workers "breaks the tie" between the choice of an exclusively skilled or an exclusively unskilled immigrant flow – and it breaks the tie in favor of skilled immigrants. A very negative elasticity of factor price for skilled workers suggests that skilled workers are highly complementary with other factors of production, particularly capital. The complementarity between native-owned capital and skills provides an economic rationale for admitting skilled workers.

This conclusion, of course, may change if the native work force is predominantly skilled. There then exist two sets of conflicting incentives. On the one hand, the immigration surplus is larger if the host country admits immigrants who most complement the skilled native workers, or unskilled immigrants. On the other hand, the immigration surplus is larger if the host country admits immigrants who most complement the native-owned capital, or skilled immigrants.

Finally, comparing Eqs. (13) and (20) yields

$$\left.\frac{\Delta Q_N}{Q}\right|_{dK=0} - \left.\frac{\Delta Q_N}{Q}\right|_{dI=0} = -\frac{1}{2c_{KK}}\left(\alpha_S c_{SK}\frac{\beta}{p_S} + \alpha_U c_{UK}\frac{1-\beta}{p_U}\right)^2 m^2 > 0, \tag{23}$$

so that the immigration surplus is larger if the capital stock in the host country is fixed.

2.4. Simulating the impact of immigration

Borjas (1995a), Borjas et al. (1997) and Johnson (1997) have used the family of models presented above to simulate the impact of immigration on the US labor market.[20] The exercise requires information on the responsiveness of factor prices to increases in labor supply. Hamermesh's (1993) comprehensive survey of the labor demand literature reveals a great deal of uncertainty in the estimates of the relevant factor price elasticities. The

[19] The first derivative evaluated at $\beta = 1$ is $(-\alpha_S^2 c_{SS} + \alpha_S \alpha_U c_{SU})$. The inequality $(c_{SS}c_{UU} - c_{SU}^2) > 0$ implies that $(-c_{SS} - c_{UU} + 2c_{SU}) > 0$, and $(-c_{SS} + c_{SU}) > 0$. As a result, the first derivative evaluated at $\beta = 1$ is positive (since $\alpha_S > \alpha_U$). The same restrictions can be used to show that the second derivative is positive everywhere.

[20] The simulation reported here uses data drawn from Borjas et al. (1997).

simulation presented here uses the following range for the vector $(\varepsilon_{SS}, \varepsilon_{UU})$: $(-0.5,-0.3)$, $(-0.9,-0.6)$, and $(-1.5,-0.8)$. This range covers most of the elasticity estimates reported in the Hamermesh survey. The cross-elasticity ε_{SU} is set to 0.05 in all the simulations. Because the weighted average of factor price elasticities is zero, these assumptions determine all the other elasticities in the model. The assumption that the wage of skilled workers is more responsive to supply shifts is consistent with the evidence, and "builds in" capital-skill complementarity into the calculations. The exercise assumes that immigration increased the labor supply of the United States by 10% – roughly the fraction of the work force that is foreign-born.

The simulation requires that workers in the US labor market be aggregated into two skill classes *and* that workers within each of the skill classes be perfect substitutes. Following Borjas et al. (1997), the exercise uses two alternative aggregations. First, all workers who are high school dropouts are defined to be in the unskilled group, while high school graduates make up the skilled group. Using this aggregation scheme, data from the 1995 Current Population Survey (CPS) then indicate that $p_S = 0.91$, but that $\beta = 0.68$. If labor's share of income is 0.7, the CPS data on the relative earnings of high school dropouts implies that the share of income accruing to skilled workers is 0.661, and that accruing to unskilled workers is 0.039.

Alternatively, divide the work force into college equivalents and high school equivalents.[21] The CPS estimates of the parameters of the skill distribution are $p_S = 0.43$ and $\beta = 0.33$; and the share of income accruing to skilled workers equals 0.371, while that accruing to unskilled workers is 0.329. Note that this aggregation of skills (unlike the one that divides the work force into high school dropouts and high school graduates) implies that the skill distribution of the immigrant work force does not differ greatly from that of the native work force.

The first two columns of Table 1 report the results using the high school dropout-graduate skill classification. If capital is perfectly inelastic, all workers lose and capital gains substantially – the income of capitalists increases by between 2.4 and 11.8%. If capital is perfectly elastic, unskilled workers lose (their earnings fall by between 1.2 and 6.1%) and skilled workers gain slightly (their earnings increase by less than 0.2%). Overall, the national income accruing to native rises by 0.1–0.4% when capital is perfectly inelastic, and by 0.1–0.2% when capital is perfectly elastic.

The last two columns of the table report the results using the high school-college equivalent aggregation. All workers still lose when capital is perfectly inelastic, and skilled workers gain and unskilled workers lose when capital is perfectly elastic. However, the losses and gains are even smaller. Immigration increases the national income accruing to natives by only 0.1–0.3% when capital is inelastic and by 0.01–0.02% when capital is elastic.

[21] The college equivalent group contains all workers who have at least a college degree, plus one-half of the workers with some college. The high school equivalent group includes workers with a high school diploma or less, plus one-half of the workers with some college. Katz and Murphy (1992) provide a detailed justification of this skill classification.

Table 1
Simulation of economic costs and benefits from immigration for the United States[a]

	Definition of skill groups			
	High school dropouts and high school graduates		High school equivalents and college equivalents	
	Capital fixed	Price of capital fixed	Capital fixed	Price of capital fixed
Assume: $(\varepsilon_{SS}, \varepsilon_{UU}) = (-0.3, -0.5)$				
Percent change in earnings of capital	2.44	–	3.71	–
Percent change in earnings of skilled workers	−0.91	0.20	−1.51	0.36
Percent change in earnings of unskilled workers	−0.28	−1.21	−1.34	−0.37
Percent change in GDP accruing to natives	0.12	0.08	0.11	0.01
Dollar gain to natives in billions, assuming $8 trillion GDP	9.76	6.65	8.94	0.91
Assume: $(\varepsilon_{SS}, \varepsilon_{UU}) = (-0.6, -0.9)$				
Percent change in earnings of capital	6.43	–	7.55	
Percent change in earnings of skilled workers	−2.29	0.46	−2.94	0.65
Percent change in earnings of unskilled workers	−3.72	−4.27	−2.89	−0.69
Percent change in GDP accruing to natives	0.27	0.14	0.22	0.02
Dollar gain to natives in billions, assuming $8 trillion GDP	24.15	10.81	17.88	1.28
Assume: $(\varepsilon_{SS}, \varepsilon_{UU}) = (-0.8, -1.5)$				
Percent change in earnings of capital	11.83	–	11.70	-
Percent change in earnings of skilled workers	−4.36	0.61	−5.08	0.92
Percent change in earnings of unskilled workers	−6.01	−6.12	−3.92	−0.98
Percent change in GDP accruing to natives	0.43	0.17	0.33	0.02
Dollar gain to natives in billions, assuming $8 trillion GDP	32.43	13.33	26.80	1.62

[a] Notes: Adapted from Borjas et al. (1997, Table 19). All simulations assume that $\varepsilon_{SU} = 0.05$; that labor's share of income is 0.7; and that the immigrant supply shock increases labor supply in the United States by 10%. The values for the other parameters are as follows. High school dropout-graduate skill grouping: $p_S = 0.91$, $\beta = 0.68$, $\alpha_S = 0.661$; $\alpha_U = 0.039$. High school-college equivalent: $p_S = 0.43$, $\beta = 0.33$, $\alpha_S = 0.371$, $\alpha_U = 0.329$.

The simulation suggests that the overall impact of immigration on the US labor market is small – regardless of how workers are grouped into different skill categories, and of the assumptions made about the factor price elasticities and the supply elasticity of capital.

3. The skills of immigrants: theory

As we have seen, the economic impact of immigration depends crucially on the differences in the skill distributions of immigrants and natives. A great deal of empirical research in economics focuses precisely on the question of how immigrant skills compare to those of native workers. Perhaps the central finding of this literature is that immigrants are not a

randomly selected sample of the population of the source countries. As a result, an understanding of the skill differentials between immigrants and natives must begin with an analysis of the factors that motivate only some persons in the source country to migrate to a particular host country.

3.1. The migration decision

It is instructive to consider a two-country model.[22] Residents of the source country (country 0) consider migrating to the host country (country 1). The migration decision is assumed to be irreversible.[23] Residents of the source country face the earnings distribution

$$\log w_0 = \mu_0 + v_0, \tag{24}$$

where w_0 gives the wage in the source country; μ_0 gives the mean earnings in the source country; and the random variable v_0 measures deviations from mean earnings and is normally distributed with mean zero and variance σ_0^2. For convenience, Eq. (24) omits the subscript that indexes a particular individual.

If the *entire* population of the source country were to migrate to the host country, this population would face the earnings distribution

$$\log w_1 = \mu_1 + v_1, \tag{25}$$

where μ_1 gives the mean earnings in the host country *for this particular population*, and the random variable v_1 is normally distributed with mean zero and variance σ_1^2. The correlation coefficient between v_0 and v_1 equals ρ_{01}.

In general, the population mean μ_1 will not equal the mean earnings of native workers in the host country. The average worker in the source country might be more or less skilled than the average worker in the host country. It is convenient to initially assume that the average person in both countries is equally skilled (or, equivalently, that any differences in average skills have been controlled for), so that μ_1 also gives the mean earnings of natives in the host country. This assumption helps isolate the impact of the selection process on the skill composition of the immigrant flow and provides a simple way for comparing the skills of immigrants and natives in the host country.

Eqs. (24) and (25) completely describe the earnings opportunities available to persons born in the source country. The insight that migration decisions are motivated mainly by wage differentials can be attributed to Sir John Hicks. In *The Theory of Wages*, Hicks (1932, p. 76) argued that "differences in net economic advantages, chiefly differences in wages, are the main causes of migration". Practically all modern studies of migration

[22] The discussion in this section is based on the presentation of Borjas (1987) and Borjas (1991).

[23] Borjas and Bratsberg (1996) generalize the model to allow for return migration by immigrants. In their model, return migration may be part of an optimal location plan over the life cycle or be induced by worse-than-expected outcomes in the host country. Regardless of the motivation, Borjas and Bratsberg show that return migration does not alter the key insights of the model, and, in fact, tends to intensify the type of selection that characterizes the immigrant flow.

decisions use this conjecture as a point of departure. Assume that the migration decision is determined by a comparison of earnings opportunities across countries, net of migration costs.[24] Define the index function

$$I = \log\left(\frac{w_1}{w_0 + C}\right) \approx (\mu_1 - \mu_0 - \pi) + (v_1 - v_0), \tag{26}$$

where C gives the level of migration costs, and π gives a "time-equivalent" measure of these costs ($\pi = C/w_0$). A person emigrates if $I > 0$, and remains in the source country otherwise.

Migration is costly, and these costs probably vary among persons – but the sign of the correlation between costs (whether in dollars on in time-equivalent terms) and wages is ambiguous. Migration costs involve direct costs (e.g., the transportation of persons and household goods), forgone earnings (e.g., the opportunity cost of a post-migration unemployment spell), and psychic costs (e.g., the disutility associated with leaving behind family ties and social networks). The distribution of the random variable π in the source country's population is

$$\pi = \mu_\pi + v_\pi, \tag{27}$$

where μ_π is the mean level of migration costs in the population, and v_π is a normally distributed random variable with mean zero and variance σ_π^2. The correlation coefficients between v_π and (v_0, v_1) are given by $(\rho_{\pi 0}, \rho_{\pi 1})$. The probability that a person migrates to the host country can be written as

$$P(z) = \Pr[v > -(\mu_1 - \mu_0 - \mu_\pi)] = 1 - \Phi(z), \tag{28}$$

where $v = v_1 - v_0 - v_\pi$, $z = -(\mu_1 - \mu_0 - \mu_\pi)/\sigma_v$, and Φ is the standard normal distribution function. Eq. (28) summarizes the economic content of the Hicksian theory of migration. In particular,

$$\frac{\partial P}{\partial \mu_0} < 0, \qquad \frac{\partial P}{\partial \mu_1} > 0, \qquad \frac{\partial P}{\partial \mu_\pi} < 0. \tag{29}$$

The emigration rate falls when the mean income in the source country rises, when the mean income in the host country falls, and when time-equivalent migration costs rise. Most studies in the literature on the internal migration of persons within a particular country focus on testing these theoretical predictions (Greenwood, 1975). The empirical evidence in these studies is generally supportive of the theory.

3.2. The self-selection of immigrants

Although it is of important to determine the size and direction of migration flows, it is

[24] The wage distributions in Eqs. (24) and (25) could be reinterpreted as giving the distributions of the present value of the earnings stream in each country. This reformulation places the model within the human capital framework proposed by Sjaastad (1962).

equally important to determine *which* persons find it most worthwhile to migrate to the host country. This question lies at the heart of the Roy model (Roy, 1951; Heckman and Honoré, 1990). Consider the conditional means $E(\log w_0 \mid \mu_0, I > 0)$ and $E(\log w_1 \mid \mu_1, I > 0)$. These conditional means give the average earnings in both the source and host countries for persons who migrate. Note that the conditional means hold μ_0 and μ_1 constant. The calculation effectively assumes that the migration flow is sufficiently small so that there are no feedback effects on the performance of immigrants (or natives) in the host country or on the performance of the "stayers" in the source country. A general equilibrium model would account for the fact that the mean of the income distributions depends on the size and composition of the immigrant flow. Because the random variables v_0, v_1, and v_π are jointly normally distributed, these conditional means are given by

$$E(\log w_0 \mid \mu_0, I > 0) = \mu_0 + \left[\frac{\sigma_0 \sigma_1}{\sigma_v} \left(\rho_{01} - \frac{\sigma_0}{\sigma_1} \right) - \rho_{\pi 0} \frac{\sigma_\pi}{\sigma_1} \right] \lambda, \tag{30}$$

$$E(\log w_1 \mid \mu_1, I > 0) = \mu_1 + \left[\frac{\sigma_0 \sigma_1}{\sigma_v} \left(\frac{\sigma_1}{\sigma_0} - \rho_{01} \right) - \rho_{\pi 1} \frac{\sigma_\pi}{\sigma_0} \right] \lambda, \tag{31}$$

where $\lambda = \phi(z)/(1 - \Phi(z))$, and ϕ is the density of the standard normal. The variable λ is inversely related to the emigration rate (Heckman, 1979, p. 156), and will be positive as long as some persons find it profitable to remain in the country of origin ($P(z) < 1$). It is easier to initially interpret the results in Eqs. (30) and (31) by assuming that $\sigma_\pi = 0$, so that time-equivalent migration costs are constant. Let $Q_0 = E(v_0 \mid \mu_0, I > 0)$ and $Q_1 = E(v_1 \mid \mu_1, I > 0)$. The Roy model identifies three cases that summarize the skill differentials between immigrants and natives:

$$Q_0 > 0 \text{ and } Q_1 > 0, \text{ if } \rho_{01} > \frac{\sigma_0}{\sigma_1} \text{ and } \frac{\sigma_1}{\sigma_0} > 1, \tag{32}$$

$$Q_0 < 0 \text{ and } Q_1 < 0, \text{ if } \rho_{01} > \frac{\sigma_1}{\sigma_0} \text{ and } \frac{\sigma_0}{\sigma_1} > 1,$$

$$Q_0 < 0 \text{ and } Q_1 > 0, \text{ if } \rho_{01} < \min\left(\frac{\sigma_1}{\sigma_0}, \frac{\sigma_0}{\sigma_1} \right).$$

Positive selection occurs when immigrants have above-average earnings in both the source and host countries ($Q_0 > 0$ and $Q_1 > 0$), and negative selection when immigrants have below-average earnings in both countries ($Q_0 < 0$ and $Q_1 < 0$). Eq. (32) shows that either type of selection requires that skills be positively correlated across countries. The variances σ_0 and σ_1 measure the "price" of skills: the greater the rewards to skills, the larger the inequality in wages.[25] Immigrants are then positively selected when the source country – *relative to the host country* – "taxes" highly skilled workers and "insures" less

skilled workers from poor labor market outcomes, and immigrants are negatively selected when the host country taxes highly skilled workers and subsidizes less skilled workers.

There exists the possibility that the host country draws persons who have below-average earnings in the source country but do well in the host country ($Q_0 < 0$ and $Q_1 > 0$). This sorting occurs when the correlation coefficient ρ_{0j} is small or negative. Borjas (1987) argues that this correlation may be negative when a source country experiences a Communist takeover. In its initial stages, this political system often redistributes incomes by confiscating the assets of relatively successful persons. Immigrants from such systems will be in the lower tail of the post-revolution income distribution, but will perform well in the host country's market economy.

Eq. (32) shows that neither differences in mean incomes across countries nor the level of migration costs determines the type of selection that characterizes the immigrant flow. Mean incomes and migration costs affect the size of the flow (and the extent to which the skills of the average immigrant differ from the mean skills of the population), but they do not determine if the immigrants are drawn mainly from the upper or lower tail of the skill distribution.

The analysis has assumed that μ_1 gives the mean income in the host country both for the average person in the source country's population as well as for the average native in the host country. The selection rules in (32) then contain all the implications of economic theory for the qualitative differences in skill distributions between immigrants and natives. Immigrants will be more skilled than natives if there is positive selection or a refugee sorting, and will be less skilled if there is negative selection. I return below to the comparison of skill distributions between immigrants and natives when mean skills differ across countries.

The discussion also assumed that migration costs are constant in the population. Eqs. (30) and (31) indicate that variable migration costs do not alter any of the selection rules if: (a) time-equivalent migration costs are uncorrelated with skills ($\rho_{\pi 0} = \rho_{\pi 1} = 0$); or (b) the ratio of variances σ_π/σ_j ($j = 0, 1$) is "small." Otherwise, variable migration costs can change the nature of selection. Suppose that π is negatively correlated with earnings, perhaps because less skilled persons find it more difficult to find jobs in the host country. This negative correlation increases the likelihood that the bracketed term in Eqs. (30) and (31) is positive, and the immigrant flow is more likely to be positively selected. Conversely, the likelihood of negative selection increases if π and earnings are positively correlated.

The theoretical analysis generates a reduced form model that describes the determinants of the relative skill composition of the immigrant flow. To simplify, suppose that time-equivalent migration costs are constant. The reduced-form equation is then given by

$$Q_1 = g(\mu_0, \mu_1, \pi, \sigma_0, \sigma_1, \rho). \tag{33}$$

[25] This interpretation of the variances follows from the definition of the log wage distribution in the host country in terms of what the population of the source country would earn if the entire population migrated there. This definition effectively holds constant the distribution of skills.

Eq. (33) summarizes the relationship between the relative skills of immigrants and the characteristics of both the source and host countries. Borjas (1987) analyzes the restrictions imposed by the income-maximization hypothesis on the function g in (33). The qualitative effects of the independent variables cannot typically be signed and can be decomposed in terms of composition effects and scale effects. A change in a variable θ might create incentives for a different type of person to migrate (a composition effect) and for a different number of persons to migrate (the scale effect).

The two effects can be isolated by estimating the two-equation structural model,

$$P = P(\mu_0, \mu_1, \pi, \sigma_0, \sigma_1, \rho), \tag{34}$$

$$Q_I = h(\sigma_0, \sigma_1, \rho)\lambda. \tag{35}$$

Eq. (34) describes the determinants of the probability of migration, and (35) describes the determinants of the relative skills of immigrants. Recall that λ is a transformation of the probability of migration. By holding λ constant, the function h in (35) nets out the scale effect and isolates the impact of source and host country characteristics on the selection of the immigrant flow.

The income-maximization hypothesis imposes the following restrictions on h, the λ-constant "immigrant quality" function:

1. an increase in σ_0 decreases the average skills of immigrant;
2. an increase in σ_1 increases the average skills of immigrants;[26]
3. an increase in ρ_{0I} increases the average skills of immigrants if there is positive selection and decreases the average skills if there is negative selection.

The Roy model generates predictions about how immigrants compare to the population of the *source* countries. This contrast is not relevant if we wish to determine the impact of immigration on the host country – that impact depends on the skill differential between immigrants and natives in the host country. The discussion introduced the immigrant-native comparison by assuming that the average person in the source country has the same skills as the average person in the host country. Different countries, however, have different skill distributions.

The skill differential between immigrants and natives in the host country, therefore, will depend both on the selection rules and on the average skill differential between the source and host countries. Suppose we interpret the mean income in the source country, μ_0, as a measure of the average skills in that country. The mean earnings of immigrants in the host country are then given by

$$\mathrm{E}(\log w_I \mid \mu_0, I > 0) = \mu_I(\mu_0) + \mathrm{E}(v_I \mid \mu_0, I > 0). \tag{36}$$

[26] An increase in σ_1 stretches the income distribution in the host country and leads to a different mean wage level in the pool of migrants even when the pool is restricted to include the same persons – so that it is not a mean-preserving shift. A simple solution to this technical detail is to define immigrant quality in terms of standardized units (or Q_I/ω_1). The prediction in the text can then be easily derived.

Eq. (36) shows that the mean income of immigrants in the host country depends on the extent to which the average skills in the source country affect earnings in the host country (i.e., $d\mu_1/d\mu_0$). If this derivative were equal to one, skills are perfectly transferable across countries, and, abstracting from selection issues, workers who originate in high-income countries would have higher earnings in the host country.

Some of the implications of the Roy model have been tested empirically by estimating the correlation between the earnings of immigrants in the United States and measures of the rate of return to skills in the source country. There exists a great deal of dispersion in skills and economic performance among immigrant groups in the United States. In 1990, immigrants originating in Mexico or Portugal had about 8 years of schooling, while those originating in Austria, India, Japan, and the United Kingdom had about 15 years. Immigrants from El Salvador or Mexico earn 40% less than natives, while immigrants from Australia or South Africa earn 30–40% more than natives.[27]

The empirical studies have typically estimated the reduced-form earnings equation in (33). The evidence provides some support for the hypothesis that immigrants originating in countries with higher rates of return to skills have lower earnings in the United States. Borjas (1987, 1991) reports that measures of income inequality in the source country, which are a very rough proxy for the rate of return to skills, tend to be negatively correlated with the earnings of immigrant men, while Cobb-Clark (1983) reports a similar finding for immigrant women.[28] Barrett (1993) shows that immigrants who enter the United States using a family reunification visa have lower earnings when they originate in countries where the income distribution has a large variance. Bratsberg (1995) documents that the foreign students who remain in the United States after completing their education earn relatively high US wages if the source country offers a low rate of return to skills, but earn low wages if the source country offers a high rate of return to skills. Finally, Taylor's (1987) case study of migration in a rural Mexican village concludes that Mexicans who migrated illegally to the United States are less skilled, on average, than the typical person residing in the village. This type of selection is consistent with the fact that Mexico has a higher rate of return to skills than the United States.[29]

[27] These statistics are reported in Borjas (1994, p. 1686).

[28] Migration decisions are typically made in a family context. Mincer's (1978) family migration model assumes that the family's objective is to maximize family income. Some persons in the household may then take actions that are not "privately" optimal (i.e., they would not have taken those actions if they wished to maximize their own individual income). The family context of immigration gives rise to "tied movers" (persons who moved, even though it was privately optimal to stay), and "tied stayers" (persons who stayed, even though it was privately optimal to move). The presence of tied movers in the immigrant flow tends to attenuate the type of selection that characterizes the immigrant population in the host country (Borjas and Bronars, 1991). The study of the economic performance of immigrant women requires a careful delineation of how the family migration decision alters the skill composition of immigrants. Such a study, however, has not yet been conducted for the United States.

3.3. Selection in observed characteristics

It is instructive to differentiate between skills that are observed and skills that are not. For simplicity, let's assume that a worker obtains s years of schooling *prior* to the migration decision, and that this educational attainment can be observed and valued properly by employers in both countries. The earnings functions are given by

$$\log w_0 = \mu_0 + \delta_0 s + \varepsilon_0, \tag{37}$$

$$\log w_1 = \mu_1 + \delta_1 s + \varepsilon_1, \tag{38}$$

where δ_j gives the rate of return to schooling in country j, and ε_j is a random variable measuring deviations in earnings due to unobserved characteristics.[30] The random variables ε_0 and ε_1 are jointly normally distributed with mean zero, variances σ_0^2 and σ_1^2, and correlation coefficient ρ_{01}. The variance σ_j^2 now measures the price of unobserved skills in country j.

Suppose the distribution of educational attainment in the source country's population is

$$s = \mu_S + \varepsilon_S, \tag{39}$$

where ε_S is normally distributed with mean zero and variance σ_s^2. In general, the random variable ε_S is correlated with ε_0 and ε_1. For analytical convenience, suppose that ε_S is uncorrelated with the difference $(\varepsilon_1 - \varepsilon_0)$.

Assume that time-equivalent migration costs are constant. The migration rate for the population of the source country is

$$P(z^*) = \Pr[\tau > -[(\mu_1 - \mu_0) + (\delta_1 - \delta_0)\mu_s - \pi] = 1 - \Phi(z^*), \tag{40}$$

where $\tau = (\varepsilon_1 - \varepsilon_0) + (\delta_1 - \delta_0)\varepsilon_S$, and $z^* = -[(\mu_1 - \mu_0) + (\delta_1 - \delta_0)\mu_S - \pi]/\sigma_\tau$.

It is easy to show that the selection in unobserved skills follows the selection rules derived earlier in Eq. (32). The mean schooling of persons who choose to migrate is

$$\mathrm{E}(s \mid \mu_S, I > 0) = \mu_s + \frac{\sigma_s^2}{\sigma_\tau}(\delta_1 - \delta_0)\lambda. \tag{41}$$

The mean schooling of immigrants is less than or greater than the mean schooling in the source country depending on which country has a higher rate of return. Highly educated workers end up in the country that values them the most.

[29] Some empirical studies also report a strong positive correlation between the earnings of immigrants in the United States and the level of economic development in the source country, as measured by per-capita GDP (Jasso and Rosenzweig, 1986). As suggested by Eq. (36), this correlation might measure the portability of human capital across countries, with capital acquired in more developed countries being more easily transferable to the US labor market.

[30] The rate of return offered by the host country to schooling acquired in the source country might have little relation to the rate of return that the host country offers to schooling acquired in the host country. In the United States, for example, the empirical evidence suggests that schooling acquired in the pre-migration period has a lower value than schooling acquired in the United States (Borjas, 1995a; Funkhouser and Trejo, 1995).

Differentiating the conditional mean in (41) yields

$$\frac{\partial E(s \mid I > 0)}{\partial \mu_s} = 1 - \frac{(\delta_1 - \delta_0)^2 \sigma_s^2}{\sigma_\tau^2} \frac{\partial \lambda}{\partial z^*}. \tag{42}$$

The definition of the variance σ_τ^2 implies that $(\delta_1 - \delta_0)^2 \sigma_s^2 < \sigma_\tau^2$. It can be shown that $0 < \partial\lambda/\partial z^* < 1$ (Heckman, 1979, p. 157). Therefore,

$$0 < \frac{\partial E(s \mid I > 0)}{\partial \mu_s} < 1. \tag{43}$$

A 1-year increase in the mean education of the source country increases the mean education of persons who actually migrate to the host country, but by less than one year.[31] The inequality in (43) implies that the variance in mean education across immigrant groups who originate in different countries but live in the same host country is smaller than the variance in mean education across the different source countries. As a result of immigrant self-selection, relatively similar persons tend to migrate to the host country. The selection process thus serves as a pre-arrival "melting pot" that makes the immigrant population in the host country more homogeneous than the population of the various countries of origin.

Superficially, it seems as if the selection rule for observable skills implicit in Eq. (41) has little to do with the selection rules for unobserved skills in (32). However, the fundamentals that drive immigrant selection are exactly the same. The sorting in observed characteristics is guided by the prices δ_0 and δ_1. The selection in unobserved characteristics is also guided by their prices, the variances σ_0^2 and σ_1^2.[32]

4. The skills of immigrants: empirics

Much of the empirical research in the immigration literature analyzes the differences in the skill distributions of immigrants and natives. Beginning with the work of Chiswick (1978) and Carliner (1980), these studies attempt to measure both the skill differential at the time of entry and how this differential changes over time as immigrants adapt to the host country's labor market. A key result of this literature is that there exists a positive correlation between the earnings of immigrants and the number of years that have elapsed since immigration.[33] As will be seen below, there has been a great deal of debate over the interpretation of this correlation.

[31] Suppose, for example, that $(\delta_1 - \delta_0) > 0$. An increase in μ_s makes it worthwhile for more persons to migrate and dilutes the mean education of the immigrant sample. Hence the increase in the conditional expectation of schooling is smaller than the increase in the population mean.

[32] Borjas et al. (1992) generalize the Roy model to show that the skill sorting of workers among n potential regions is also guided by the regional distribution of the returns to skills. The n-country model is difficult to solve (and estimate) unless one makes a number of simplifying assumptions about the joint distribution of skills. Dahl (1997) provides a good discussion of the challenges encountered in estimating polychotomous choice models in the context of internal migration decisions.

4.1. The identification problem

The empirical analysis of the relative economic performance of immigrants was initially based on the cross-section regression model:

$$\log w_i = X_i\beta_0 + \beta_1 I_i + \beta_2 y_i + \varepsilon_i, \tag{44}$$

where w_i is the wage rate of person i in the host country; X_i is a vector of socioeconomic characteristics (often including age and education); I_i is a dummy variable set to unity if person i is foreign-born; and y_i gives the number of years that the immigrant has resided in the United States and is set to zero if i is a native.[34] Because the vector X controls for age, the coefficient β_2 measures the differential value that the host country's labor market attaches to time spent in the host country versus time spent in the source country.

Beginning with Chiswick (1978), cross-section studies of immigrant earnings have typically found that β_1 is negative and β_2 is positive. Chiswick's analysis of the 1970 US Census data indicates that immigrants earn about 17% less than "comparable" natives at the time of entry, and this gap narrows by slightly over 1 percentage point per year.[35] As a result, immigrant earnings overtake those of their native counterparts after about 15 years in the United States. The steeper age earnings profiles of immigrants was interpreted as saying that immigrants accumulated human capital – relative to natives – as the "Americanization" process took hold, closing the wage gap between the two groups. The overtaking phenomenon was then explained in terms of a selection argument: immigrants are "more able and more highly motivated" than natives (Chiswick (1978, p. 900), or immigrants "choose to work longer and harder than nonmigrants" (Carliner, 1980, p. 89). As we have seen, these assumptions about the selection process are not necessarily implied by income-maximizing behavior on the part of immigrants.

Borjas (1985) suggested an alternative interpretation of the cross-section evidence. Instead of interpreting the positive β_2 as a measure of assimilation, he argued that the cross-section data might be revealing a decline in relative skills across successive immigrant cohorts.[36] In the United States, the postwar era witnessed major changes in immigration policy and in the size and national origin mix of the immigrant flow. If these changes generated a less-skilled immigrant flow, the cross-section correlation indicating that more recent immigrants earn less may say little about the process of wage convergence, but may instead reflect innate differences in ability or skills across cohorts.[37]

[33] Although most of the empirical evidence focuses on the US experience, the literature also suggests that this correlation is observed in Canada (Bloom and Gunderson, 1991; Baker and Benjamin, 1994), Australia (Beggs and Chapman, 1991), and Germany (Dustmann, 1993; Pischke, 1993).

[34] The models actually used in empirical studies typically include higher-order polynomials in age and years-since-migration. These non-linearities, however, do not affect the key identification issue.

[35] Chiswick's (1978) study uses log annual earnings as the dependent variable and includes education, potential experience (and its squared), the log of weeks worked, and some regional characteristics in the vector X.

[36] Douglas (1919) presents a related discussion of cohort effects in the context of early 20th century immigration.

The identification of aging and cohort effects raises difficult methodological problems in many demographic contexts. Identification requires the availability of longitudinal data where a particular worker is tracked over time, or, equivalently, the availability of a number of randomly drawn cross-sections so that specific cohorts can be tracked across survey years. Suppose that a total of Ω cross-section surveys are available, with cross-section τ ($\tau = 1, \ldots, \Omega$) being obtained in calendar year T_τ. Pool the data for immigrants and natives across the cross-sections, and consider the regression model
Immigrant equation:

$$\log w_{\ell\tau} = X_{\ell\tau}\phi_{i\tau} + \delta_i A_{\ell\tau} + \alpha y_{\ell\tau} + \beta C_{\ell\tau} + \sum_{\tau=1}^{\Omega} \gamma_{i\tau}\pi_{\ell\tau} + \varepsilon_{\ell\tau}, \tag{45}$$

Native equation:

$$\log w_{\ell\tau} = X_{\ell\tau}\phi_{n\tau} + \delta_n A_{\ell\tau} + \sum_{\tau=1}^{\Omega} \gamma_{n\tau}\pi_{\ell\tau} + \varepsilon_{\ell\tau}, \tag{46}$$

where $w_{l\tau}$ gives the wage of person l in cross-section τ; X gives a vector of socioeconomic characteristics; A gives the worker's age at the time the cross-section survey is observed; $C_{l\tau}$ gives the calendar year in which the immigrant arrived in the host country; $y_{l\tau}$ gives the number of years that the immigrant has resided in the host country ($y_{l\tau} = T_\tau - C_{l\tau}$); and $\pi_{l\tau}$ is a dummy variable indicating if person l was drawn from cross-section τ.[38]

Because the worker's age is a regressor, the coefficient α measures the differential value of a year spent in the host country versus a year spent in the source country. Define

$$\alpha^* = \left.\frac{\partial \log w_l}{\partial t}\right|_{\text{Immigrant}} - \left.\frac{\partial \log w_l}{\partial t}\right|_{\text{Native}} = (\delta_i + \alpha) - \delta_n, \tag{47}$$

where the derivatives account for the fact that both age and the number of years-since-migration change over time. The parameter α^* measures the rate of wage convergence between immigrants and natives (an aging effect); the coefficient β indicates how the earnings of immigrants are changing across cohorts, and measures the cohort effect, and the vectors γ_i and γ_n give the impact of aggregate economic conditions on immigrant and natives wages, respectively, and measure period effects.

The identification problem arises from the identity

$$y_{l\tau} \equiv \sum_{\tau=1}^{\Omega} \pi_\tau (T_\tau - C_{l\tau}). \tag{48}$$

[37] Endogenous return migration can also generate skill differentials among immigrant cohorts. Suppose, for example, that return migrants have relatively lower wages. Earlier cohorts will then have higher average wages than more recent cohorts.

[38] A more general model would allow for non-linearities in the age, years-since-migration, and year-of-arrival variables, variation in the coefficient vector (ϕ, δ) over time, as well as differences in the coefficient α across immigrant cohorts. For the most part, these generalizations do not affect the discussion of identification issues.

Eq. (48) introduces perfect collinearity among the variables y_{it}, C_{it} and π_{it} in the immigrant earnings function. As a result, the key parameters of interest – α, β, and the vector γ_i – are not identified. Some type of restriction must be imposed if we wish to separately identify the aging effect, the cohort effect, and the period effects. Borjas (1985) proposed the restriction that the period effects are the same for immigrants and natives:

$$\gamma_{i\tau} = \gamma_{n\tau}, \qquad \forall \tau. \tag{49}$$

Put differently, trends in aggregate economic conditions change immigrant and native wages by the same percentage amount. A useful way of thinking about this restriction is that the period effects for immigrants are calculated from *outside* the immigrant wage determination system.[39]

Friedberg (1992) argued that the generic model in (45) and (46) ignores an important aspect of immigrant wage determination: the role of age-at-arrival in the host country. The US data suggest a strong negative correlation between age-at-arrival and entry earnings. The identification problem, however, does not disappear when the entry wage of immigrants depends on age-at-migration. Rather, it becomes more severe. Consider the following generalization of Eq. (45):

$$\log w_{it} = X_{it}\phi_{it} + \delta_i A_{it} + \alpha y_{it} + \beta C_{it} + \theta M_{it} + \sum_{\tau=1}^{\Omega} \gamma_{i\tau}\pi_{it} + \varepsilon_{it}, \tag{50}$$

where M_{it} gives the immigrant's age at migration. As before, the parameter vector (α, β, γ_i) in (50) cannot be identified because the identity in Eq. (48) still holds. The inclusion of the age-at-migration variable, however, introduces yet another identity: $M_{it} \equiv A_{it} - y_{it}$. Moreover, the perfect collinearity introduced by this identity remains even after the period effects are assumed to be the same for immigrants and natives. As a result, an additional restriction must be imposed on the data. One possible restriction is that the coefficient of the age variable is the same for immigrants and natives. The estimation of the system in (46) and (50) then requires that

$$\delta_i = \delta_n \qquad \text{and} \qquad \gamma_{i\tau} = \gamma_{n\tau}, \qquad \forall \tau. \tag{51}$$

The assumption that the age coefficient is the same in both the immigrant and native samples is very restrictive, and contradicts the notion of specific human capital. After all, it is very unlikely that a year of pre-migration "experience" for immigrants has the same value in the host country's labor market as a year of experience for the native population. Nevertheless, *some* restriction must be imposed if age-at-migration is to have an independent effect on the wage determination process. An alternative approach might model the age-at-migration effect as a step function: persons who migrate as children face different opportunities in the host country than those who migrate as adults. This

[39] Eq. (49) is less restrictive than it seems. After all, it does not define which native group experienced the same period effects as the immigrant population.

specification would break the perfect collinearity between age, age-at-migration, and years-since-migration.

Overall, the lesson is clear: estimates of aging and cohort effects are conditional on the imposed restrictions. Different restrictions lead to different estimates of the underlying parameters of interest.

4.2. Economic assimilation

Even after the analysis has allowed for the possibility of cohort effects, there seems to be a great deal of confusion in the empirical literature about whether immigrants in the United States experience a substantial degree of "economic assimilation."[40] Part of the confusion can be traced directly to a conceptual disagreement over the definition of assimilation.

The *Oxford English Dictionary* defines assimilation as "the action of making or becoming like," while *Webster's Collegiate Dictionary* defines it as "the process whereby individuals or groups of differing ethnic heritage are absorbed into the dominant culture of a society." Any sensible definition of economic assimilation, therefore, must define a base group that the immigrants are assimilating to. Beginning with Chiswick's (1978) study of the "Americanization" of the foreign-born in the United States, many studies implicitly or explicitly use a definition that equates the concept of economic assimilation with the rate of wage convergence between immigrants and natives in the host country. This definition of economic assimilation is given by α^* in Eq. (47).

LaLonde and Topel (1992, p. 75) propose a very different definition of the process: "assimilation occurs if, between two observationally equivalent (foreign-born) persons, the one with greater time in the United States typically earns more" (LaLonde and Topel, 1992, p. 75). In terms of the econometric model in Eqs. (45) and (46), the LaLonde–Topel definition is simply the parameter α, the coefficient of years-since-migration in the immigrant earnings function.

The two alternative definitions of economic assimilation, α^* and α, stress different concepts and address different questions. The parameter α defines assimilation by comparing the economic value (in terms of the host country's labor market) of a year spent in the host country *relative* to a year spent in the source country. Hence the base group in the LaLonde–Topel definition of economic assimilation is *the immigrant himself*. Immigrants assimilate in the sense that they are picking up skills in the host country's labor market that they would not be picking up if they remained in the source country.

A positive α, however, provides no information whatsoever about the trend in the economic performance of immigrants in the host country – relative to that of natives. Suppose, for example, that the coefficient of the age variable in the immigrant earnings function is smaller than the respective coefficient in the native earnings function ($\delta_i < \delta_n$).[41] It is then numerically possible to estimate a very positive α, conclude that

[40] The confusion is also present in the empirical studies of the Canadian experience. See, for example, Bloom and Gunderson (1991), Baker and Benjamin (1994) and Bloom et al. (1995).

there is economic assimilation in the LaLonde–Topel sense, and observe that immigrant earnings keep falling further behind those of natives over time ($\alpha^* < 0$).

The ambiguities introduced by the choice of a base group pervade studies of immigrant economic performance. For example, the discussion of identification issues ignored the question of exactly which variables should enter the standardizing vector X in the earnings functions (see Eqs. (45) and (46)). The choice of standardizing variables is not discussed seriously in most empirical studies in labor economics, where the inclusion criteria seems to be determined by the list of variables available in the survey data under analysis. But this issue plays a significant role in the study of immigrant wage determination. The disagreement in the empirical literature over the relative economic status of immigrants in the United States arises not only because different studies use different definitions of economic assimilation, but also because different studies use different standardizing variables. As a result, the base group differs haphazardly from study to study.

For example, many studies include a worker's educational attainment (measured as of the time of the survey) in the vector X, so that the cohort and aging effects are measured relative to native workers who have the same schooling. This standardization introduces two distinct problems. First, part of the adaptation process experienced by immigrants might include the acquisition of additional schooling. By controlling for schooling observed at the time of the survey, the analysis hides the fact that there might be a great deal of wage convergence between immigrants and natives. Second, the inclusion of schooling in the earnings functions introduces the possibility of "over-controlling" – of addressing such narrow questions that the empirical evidence has little economic or policy significance. It might be interesting to know that the wage of an immigrant high school dropout converges to that of a native high school dropout, but it is probably more important to determine how the skills of the immigrant high school dropout compare to those of the typical native worker. After all, economic theory teaches us that the economic impact of immigration depends on how immigrants compare to natives, and *not* on how immigrants compare to statistically similar natives.

4.3. Empirical evidence for the United States

A large literature summarizes the trends in the skills and wages of immigrants in the United States.[42] Almost all of these studies combine data from various US Census cross-sections to identify the aging and cohort effects. The essence of the empirical evidence reported in this literature can be obtained by estimating the following regression model in the sample of working men in each Census cross-section:[43]

[41] This is not an idle speculation. Most empirical studies for the United States do, in fact, show that the age coefficient in the immigrant regression is much smaller than the respective coefficient in the native regression; see Borjas (1995a), LaLonde and Topel (1992), and Funkhouser and Trejo (1995). Baker and Benjamin (1994) also find the same difference in the age coefficients in the Canadian context.

[42] See, for example, Borjas (1985, 1995a), Chiswick (1978, 1986), Duleep and Regets (1997), Funkhouser and Trejo (1995), LaLonde and Topel (1992), National Research Council (1997, Chapter 5), and Yuengert (1994).

$$\log w_{I\tau} = X_{I\tau}\beta_\tau + \delta_\tau I_{I\tau} + \varepsilon_{I\tau}, \tag{52}$$

where $w_{I\tau}$ is the wage of person I in the cross-section observed at time τ ($\tau = 1960, 1970, 1980, 1990$); X is a vector of socioeconomic characteristics; and $I_{I\tau}$ is a dummy variable set to unity if person I is an immigrant and zero otherwise. The coefficient δ_τ gives the log wage differential between immigrants and natives at time τ. The analysis uses two alternative specifications of the vector X. In the first, this vector contains only an intercept. In the second, X includes the worker's educational attainment, a fourth-order polynomial in the worker's age, and variables indicating the Census region of residence.[44]

The first row of Table 2 summarizes the trend in the relative wage of immigrant men. The sign and magnitude of the unadjusted wage differential between immigrant and native men changed substantially between 1960 and 1990. In 1960, immigrants earned about 4% more than natives did; by 1990, immigrants earned 16.3% less. About half of the decline in the relative wage of immigrants can be explained by changes in observable socioeconomic characteristics, particularly educational attainment.

The second row of the table documents the trend in the relative wage of "new" immigrants (these immigrants have been in the United States for less than 5 years as of the time of the Census).[45] The latest cohort of immigrants earned 13.9% less than natives in 1960 and 38.0% less in 1990. A substantial fraction of the decline in the relative wage of new immigrants can also be explained by changes in observable socioeconomic characteristics.

As indicated earlier, the *interpretation* of these trends requires that restrictions be imposed on the period effects. If changes in aggregate economic conditions did not affect the relative wage of immigrants (as implied by Eq. (49)), the cohort effects in Table 2 then indicate that the relative skills of immigrants declined across successive immigrant cohorts.[46] This interpretation, therefore, uses a difference-in-differences estimator to identify the trend in relative immigrant skills.[47]

The remaining rows of Table 2 show how the relative wage of a particular immigrant cohort changes over time. These statistics are obtained by estimating the regression model in (52) on a pooled sample that includes natives in a particular age group and immigrants who arrived at a particular point in time and are in the same age group. For example, the

[43] The empirical analysis reported below uses the sample of men aged 25–64 who are employed in the civilian sector, are not self-employed, and do not live in group quarters.

[44] The vector of educational attainment indicates if the worker has less than 9 years of schooling; 9–11 years; 12 years; 13–15 years; and 16 or more years. The Census region of residence dummies indicate if the worker lives in the Northeast region, the North Central region, the South region, or the West region.

[45] The year-of-migration question in the 1960 Census differs from that in the post-1960 Censuses. In 1960, persons reported where they lived 5 years ago. The new immigrant cohort in 1960 is composed of persons who are either naturalized citizens or non-citizens, and were residing abroad in 1955. Since 1970, persons are asked when they came to the United States to stay, and the new immigrant cohorts in these Censuses are composed of persons who are either naturalized citizens or non-citizens, and who came to the United States "to stay" in the last 5 years. Finally, the 1955–1960 cohort can be defined uniquely only in the 1960 and 1970 Censuses.

[46] The implicit link between wages and skills, of course, presupposes that the data are being interpreted through the lens of a human capital model of wage determination.

Table 2
Relative wage of immigrant men in the United States, 1960–1990[a]

	Unadjusted relative wage				Adjusted relative wage			
Group	1960	1970	1980	1990	1960	1970	1980	1990
All immigrants	0.041	−0.001	−0.097	−0.163	0.013	−0.017	−0.071	−0.100
	(0.005)	(0.005)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)
Newly arrived immigrants	−0.139	−0.188	−0.328	−0.380	−0.162	−0.198	−0.241	−0.269
	(0.014)	(0.011)	(0.008)	(0.007)	(0.013)	(0.010)	(0.008)	(0.006)
1955–1960 arrivals								
25–34 in 1960	−0.094	0.062	–	–	−0.128	0.049	–	–
	(0.019)	(0.019)			(0.018)	(0.018)		
35–44 in 1960	−0.140	−0.010	–	–	−0.181	−0.012	–	–
	(0.025)	(0.027)			(0.023)	(0.025)		
45–54 in 1960	−0.172	−0.056	–	–	−0.218	−0.097	–	–
	(0.036)	(0.039)			(0.033)	(0.036)		
1965–1970 arrivals								
15–24 in 1970	–	–	−0.047	−0.067	–	–	0.023	0.032
			(0.015)	(0.016)			(0.015)	(0.015)
25–34 in 1970	–	−0.139	−0.061	−0.022	–	−0.173	−0.046	−0.014
		(0.014)	(0.015)	(0.016)		(0.014)	(0.014)	(0.015)
35–44 in 1970	–	−0.170	−0.159	−0.087	–	−0.190	−0.121	−0.052
		(0.019)	(0.021)	(0.026)		(0.017)	(0.020)	(0.024)
45–54 in 1970	–	−0.248	−0.247	–	–	−0.220	−0.194	–
		(0.029)	(0.034)			(0.026)	(0.032)	
1975–1980 arrivals								
25–34 in 1980	–	–	−0.244	−0.164	–	–	−0.200	−0.087
			(0.010)	(0.011)			(0.010)	(0.011)
35–44 in 1980	–	–	−0.295	−0.271	–	–	−0.285	−0.213
			(0.016)	(0.019)			(0.016)	(0.017)
45–54 in 1980	–	–	−0.353	−0.302	–	–	−0.337	−0.277
			(0.026)	(0.033)			(0.016)	(0.031)

[a] Notes: Standard errors are reported in parentheses. The adjusted relative wage is obtained from a regression that includes a fourth-order polynomial in age, a vector of dummy variables indicating the worker's educational attainment, and a vector of dummy variables indicating the region of residence. The statistics are calculated in the sample of men aged 25–64 (unless otherwise indicated), who work in the civilian sector, who are not self-employed, and who do not reside in group quarters.

third row of the table report the results from regressions that includes natives aged 25–34 as of 1960 and immigrants who were also 25–34 as of 1960 and arrived between 1955 and 1960. This sample is then "tracked" across Censuses. The wage of these immigrants not only caught up with, but also overtook, the wage of similarly aged natives; an initial 9.4% wage disadvantage in 1960 became a 6.2% wage advantage by 1970. The post-1965 immigrants, however, generally start with a larger wage disadvantage and have a smaller rate of relative wage growth.

Although much of the empirical literature focuses on the secular trend in the mean of the relative wage of immigrants, it is useful to describe the evolution of the income distributions of immigrants and natives (Butcher and DiNardo, 1996). A simple representation of these trends can be obtained by using each Census cross-section to estimate the following regression in the sample of *native* workers:

$$\log w_{l\tau} = X_{l\tau}\beta_\tau + \varepsilon_{l\tau}. \tag{53}$$

The residuals from each regression are used to divide the native wage distribution into deciles, with $v_{k\tau}$ giving the benchmark for the kth decile in Census year τ (with $v_{0\tau} = -\infty$, and $v_{10,\tau} = +\infty$). By construction, 10% of the native sample lies in each decile. As before, the analysis uses two alternative specifications of X. The first includes only an intercept; the second includes educational attainment, age, and region of residence.

To calculate how many immigrants place in each decile of the native wage distribution, we can use the equations estimated in (53) to predict the residuals for the immigrant sample in each cross-section. Let $\tilde{v}_{l\tau}$ be the residual for immigrant l in year τ and define

$$d_{k\tau} = \Pr[v_{k-1,\tau} < \tilde{v}_{l\tau} < v_{k\tau}]. \tag{54}$$

The statistic $d_{k\tau}$ gives the fraction of the immigrant sample that lies in the *kth* decile of the native wage distribution in year τ.

The top panel of Table 3 reports the calculations for the immigrant sample, while the bottom panel reports the distributions for the sample of newly arrived immigrants (where the calculation in Eq. (54) uses only the sample of immigrants who have been in the United States less than 5 years).[48] The 1960–1990 period witnessed a substantial change in the *relative* wage distribution of immigrants. In 1960, 17.4% of all immigrants and 28.5% of new immigrants fell in the bottom two deciles of the native wage distribution. By 1990,

[47] However, the US wage structure changed markedly in the 1980s (Murphy and Welch, 1992; Katz and Murphy, 1992), with a substantial decline in the relative wage of less-skilled workers. As a result, the assumption that the period effects are the same for immigrants and natives is probably invalid. Borjas (1995a) presents some evidence suggesting that the changes in the US wage structure were not sufficiently large to account for the cohort effects reported in Table 2.

[48] This methodology can also be used to describe how the wage distribution of a particular immigrant cohort evolves over time and to compare this evolution to that experienced by native workers. This type of analysis would allow the calculation of rates of "distributional convergence." The results (not shown) suggest that the 1955–1960 cohort experienced substantial distributional convergence, but that this type of convergence is rarer for the post-1965 cohorts.

Table 3
Immigrant placement in the US native wage distribution, by decile[a]

Decile of native distribution	Unadjusted distribution				Adjusted distribution			
	1960	1970	1980	1990	1960	1970	1980	1990
All immigrants								
1	7.7	11.2	15.4	18.3	9.9	12.1	14.3	15.1
2	9.7	10.3	13.1	14.6	9.9	10.6	12.8	13.4
3	12.3	10.4	11.3	10.6	9.9	9.9	11.2	11.4
4	9.2	10.0	9.6	9.5	9.7	9.4	9.6	9.7
5	10.8	9.2	8.7	8.9	9.4	8.6	8.9	8.9
6	9.6	10.5	8.4	7.5	9.9	9.7	8.3	8.2
7	9.7	8.0	7.2	6.5	10.5	9.5	8.2	7.9
8	9.7	9.5	7.6	7.0	9.9	9.4	8.1	7.8
9	10.6	10.0	8.1	8.1	10.0	10.0	8.2	7.9
10	10.9	11.0	10.5	8.9	10.8	10.7	10.4	9.7
Newly arrived immigrants								
1	14.6	19.8	26.9	30.0	18.5	22.3	23.5	24.5
2	13.9	15.8	18.1	18.9	12.6	14.5	17.1	17.5
3	15.6	11.6	13.1	10.8	12.7	11.0	12.2	12.2
4	8.9	9.3	8.7	8.4	8.8	8.9	9.1	9.1
5	8.7	7.3	6.7	6.9	8.5	7.2	7.4	7.1
6	7.3	7.5	5.5	4.7	8.1	7.9	5.7	5.9
7	7.2	5.6	4.3	4.0	7.3	6.9	5.3	5.4
8	7.8	6.9	4.3	4.2	8.0	6.2	5.1	5.1
9	7.1	7.7	4.2	5.0	6.9	6.7	5.2	4.9
10	8.8	8.6	8.2	7.0	8.6	8.4	9.2	8.4

[a] Notes: The adjusted distributions are obtained from a regression that includes a fourth-order polynomial in age, a vector of dummy variables indicating the worker's educational attainment, and a vector of dummy variables indicating the region of residence. The statistics are calculated in the sample of men aged 25–64 who work in the civilian sector, who are not self-employed, and who do not reside in group quarters.

32.9% of all immigrants and 48.9% of new immigrants fell in the bottom two deciles. Put differently, the decline in the average relative wage of successive immigrant cohorts can be attributed to the increasing likelihood that new immigrants fall into the very bottom of the native wage distribution.[49]

Finally, it is instructive to estimate the regression model presented in the previous section in Eqs. (45) and (46) to illustrate the importance of choosing a definition of economic assimilation. The regression results reported in Table 4 are drawn from Borjas (1995a), pool data from the 1970, 1980, and 1990 Censuses, and include third-order

[49] The results presented in Table 3 are consistent with the evidence presented by Borjas et al. (1997, Table 15) and Card (1997, Table 2). Butcher and DiNardo (1996) use a kernel density estimator and find that the differences between the wage distributions of immigrants and natives have not changed much in the past three decades. The Butcher–DiNardo analysis, however, controls for differences in educational attainment among the various groups.

Table 4
Log wage regressions estimating aging and cohort effects in the United States[a]

Variable	Model			
	(1)		(2)	
	Native	Immigrant	Native	Immigrant
Intercept	−0.624	−0.971	−1.222	−1.057
	(0.057)	(0.062)	(0.054)	(0.059)
Age at time of survey	0.118	0.129	0.094	0.088
	(0.004)	(0.005)	(0.004)	(0.004)
Age squared	−0.002	−0.002	−0.002	−0.002
	(0.000)	(0.000)	(0.000)	(0.000)
Age cubed × 10^{-4}	0.104	0.145	0.074	0.086
	(0.008)	(0.008)	(0.007)	(0.008)
Educational attainment at time of survey	–	–	0.060	0.047
			(0.000)	(0.000)
Years since migration at time of survey	–	0.011	–	0.019
		(0.001)		(0.001)
Years since migration squared	–	0.000	–	0.000
		(0.000)		(0.000)
Years since migration cubed × 10^{-4}	–	0.004	–	0.032
		(0.004)		(0.004)
Cohort effects: relative to 1985 –1989 arrivals				
Arrived in 1980 –1985	–	0.000	–	0.004
		(0.005)		(0.005)
Arrived in 1975 –1979	–	0.061	–	0.059
		(0.005)		(0.005)
Arrived in 1970 –1974	–	0.097	–	0.095
		(0.007)		(0.007)
Arrived in 1965 –1969	–	0.153	–	0.113
		(0.008)		(0.008)
Arrived in 1960 –1964	–	0.202	–	0.137
		(0.010)		(0.010)
Arrived in 1950 –1959	–	0.235	–	0.160
		(0.012)		(0.012)
Arrived prior to 1950	–	0.235	–	0.146
		(0.016)		(0.017)
Period effects: relative to 1990 observation				
Observation drawn from 1970 Census	0.007	0.007	0.025	0.025
	(0.008)	(0.008)	(0.011)	(0.011)
Observation drawn from 1980 Census	0.048	0.048	−0.001	−0.001
	(0.006)	(0.006)	(0.008)	(0.008)
Estimated assimilation over first 10 years				
Using α^*	0.060		0.076	
Using α	0.099		0.149	
Estimated assimilation over first 20 years				
Using α^*	0.076		0.100	
Using α	0.175		0.235	

[a] Notes: Adapted from Borjas (1995a, Table 5). Standard errors are reported in parentheses. The regressions are estimated in the sample of men aged 25–64, who work in the civilian sector, who are not self-employed, and who do not reside in group quarters, and use the 1970, 1980, and 1990 Census cross-sections. Model (2) also includes a dummy variable indicating if the worker lives in a metropolitan area.

polynomials in age and years-since-migration.[50] The bottom rows of the table use the two alternative definitions of economic assimilation (α^* and α) to calculate the extent of economic assimilation experienced either during the first 10 or first 20 years in the United States.

The regression results reported in column (1) show that the wage of immigrants – *relative to natives* – increases by 6.0 percentage points during the first 10 years in the United States and by 9.9 points during the first 20 years. The LaLonde–Topel definition of assimilation, however, suggests that the wage of immigrants rises by 7.6 percentage points in the first 10 years and by 14.9 points in the first 20 years. The regression in column (2) includes educational attainment as a regressor and the rate of economic assimilation increases. In other words, immigrants experience greater economic assimilation relative to workers who have the same schooling. In view of the huge variation in the rates of "economic assimilation" estimated from the *same* regression model, it is not too surprising that the empirical literature disagrees over how much economic progress immigrants experience in the United States.

4.4. Convergence and conditional convergence

The confusion over the measurement of economic assimilation has motivated some researchers to estimate more directly the correlation between the skills of immigrants at the time of entry and the post-migration rate of human capital acquisition (Duleep and Regets, 1996, 1997; Borjas, 1999). A simple two-period model of the human capital accumulation process provides a way of thinking about this correlation.[51] Let K give the number of efficiency units that an immigrant has acquired in the source country. Because human capital may be partly specific, a fraction δ of these efficiency units evaporate when the worker emigrates. The number of effective efficiency units that the immigrant can rent out in the host country is $E = (1 - \delta)K$.

An immigrant lives for two periods in the host country. During the investment period, the immigrant devotes a fraction q of his human capital to the production of additional human capital, and this investment increases the number of available efficiency units in the payoff period by $g \times 100\%$. If the market-determined rental rate for an efficiency unit in the host country is one dollar, the present value of the post-migration income stream is

$$V = (1 - \delta)K(1 - q) + \rho[(1 - \delta)K(1 + g)], \tag{55}$$

where ρ is the discounting factor.[52]

The human capital production function is given by

[50] The regression models estimated in Table 4 also allow the coefficients for the linear term in age and years of schooling to vary over time; see Borjas (1995a) for additional details. The age and schooling coefficients reported in the table are those referring to the 1990 Census.

[51] See Borjas (1999) for a detailed discussion of this framework. A more general theory would model jointly both the human capital investment decision and the decision to emigrate the source country.

[52] The parameter ρ depends on the immigrant's discount rate and on the probability that the immigrant will stay in the host country (and collect the returns on the investments that are partly specific to the host country).

$$gE = (qE)^{\alpha}E^{\beta}, \tag{56}$$

where $\alpha < 1$. Immigrants with higher levels of human capital at the time of entry may be more efficient at acquiring additional human capital. This complementarity between "pre-existing" skills and the skills acquired in the post-migration period suggests that β is positive. However, because the costs of human capital investments are mostly forgone earnings, higher initial skills may make it very expensive to acquire additional skills. This "substitutability" would suggest that β is negative.

Ben-Porath's (1967) neutrality assumption states that these two effects exactly offset each other and β is zero, so that the marginal cost curve of producing human capital is independent of the worker's initial stock. Hence the *dollar* age-earnings profiles of workers who differ only in their initial stock of human capital are parallel to each other. Most empirical studies of earnings determination analyze the characteristics of *log* age-earnings profiles. Hence it is analytically convenient to define a different type of neutrality. Rewrite the human capital production function as:

$$g = q^{\alpha}E^{\alpha+\beta-1}. \tag{57}$$

Eq. (57) relates the rate of human capital accumulation (g) to the fraction of efficiency units used for investment purposes (q). Define "relative neutrality" as the case where the rate of human capital accumulation is independent of the initial level of effective capital, so $\alpha + \beta = 1$. If $\alpha + \beta > 1$, the rate of human capital accumulation is positively related to initial skills, and we have "relative complementarity." If $\alpha + \beta < 1$, the rate of human capital accumulation is negatively related to initial skills, and we have "relative substitutability."

Immigrants choose the rate of human capital accumulation that maximizes the post-migration present value of earnings. The optimal level of investment is

$$q = (\alpha\rho)^{1/(1-\alpha)}E^{(\alpha+\beta-1)/(1-\alpha)}. \tag{58}$$

If there is relative complementarity, highly skilled workers invest more; if there is relative substitutability, the more skilled invest less.

Let Δ be the percentage wage growth experienced by an immigrant in the host country:

$$\Delta = \frac{(1-\delta)K(1+g) - (1-\delta)K(1-q)}{E} = g + q. \tag{59}$$

The relationship between initial skills and wage growth is

$$\frac{d\Delta}{dE} = (\alpha + \beta - 1)\frac{(1+\alpha\rho)q}{\alpha\rho(1-\alpha)E}. \tag{60}$$

The log wage at the time of entry is

$$\log w_0 = \log E + \log(1-q), \tag{61}$$

and the relationship between the entry wage and initial skills is

$$\frac{d\log w_0}{dE} = \frac{1}{E}\left[1 - \frac{q}{1-q}\,\frac{\alpha+\beta-1}{1-\alpha}\right]. \tag{62}$$

The positive sign of the first term inside the brackets of (62) suggests that higher initial skills increase entry wages simply because those skills are valued by the host country's employers. Skills at the time of entry, however, also affect the investment rate. Define κ^* as

$$\kappa^* = \frac{(1-q)(1-\alpha)}{q} > 0. \tag{63}$$

By definition, the log entry wage is independent of the initial endowment of human capital when $\alpha + \beta - 1 = \kappa^*$. The inspection of Eqs. (60) and (62) reveal four cases that summarize the potential relationship between the log entry wage and the rate of wage growth:

1. Relative substitution between pre- and post-migration human capital ($\alpha + \beta - 1 < 0$). Skilled immigrants invest less, earn more at the time of entry, and experience less wage growth. There is a negative correlation between log entry wages and the rate of wage growth.
2. Relative neutrality in the human capital production function ($\alpha + \beta - 1 = 0$). Skilled immigrants devote the same fraction of time to human capital investments as less skilled immigrants, but earn more. There is zero correlation between log entry wages and wage growth.
3. Weak relative complementarity in human capital ($0 < \alpha + \beta - 1 < \kappa^*$). Skilled immigrants invest more, and Eq. (62) indicates that these immigrants also have higher entry wages. There is a positive correlation between log entry wages and wage growth.
4. Strong relative complementarity in human capital ($0 < \kappa^* < \alpha + \beta - 1$). The rate of human capital investment is so high for skilled workers that they actually earn less initially. There is a negative correlation between log entry wages and wage growth.[53]

These cases summarize the implications of human capital theory for the *unconditional* correlation between entry wages and the rate of wage growth. It is also of interest to determine the sign of the *conditional* correlation between log entry wages and the rate of wage growth. This conditional correlation holds initial skills constant. Differences in discounting factors (ρ) generate differences in entry wages and wage growth among immigrants. It is easy to show that

$$\left.\frac{d\log w_0}{d\rho}\right|_E = \frac{-1}{1-q}\,\frac{dq}{d\rho} < 0, \tag{64}$$

[53] A fifth case, where $\alpha + \beta - 1 = \kappa^*$, is also possible. In this case, skilled immigrants invest more but entry wages are independent of the level of effective human capital.

$$\left.\frac{d\Delta}{d\rho}\right|_E = \frac{dq}{d\rho}\left(1 + \frac{1}{\rho}\right) > 0, \tag{65}$$

since $dq/d\rho > 0$. Eqs. (64) and (65) indicate a negative correlation between the log entry wage of immigrants and the rate of wage growth, holding initial skills constant. In other words, the theory predicts "conditional convergence."[54]

One can calculate the correlation between the rate of wage growth and the log entry wages in the host country by tracking specific immigrant cohorts over time. Consider the cohort of immigrants who migrated from country j at time t, when they were k years old. Their log wage at the time of entry is given by $w_{jk}(t)$. The rate of wage growth of this immigrant cohort over the (t, t') time interval is

$$\Delta w_{jk}(t, t') = [w_{jk}(t') - w_{jk}(t)]. \tag{66}$$

Consider the regression model:

$$\Delta w_{jk}(t, t') = \theta w_{jk}(t) + \xi_{kt} + \nu_{jk}, \tag{67}$$

where ξ_{kt} gives a year-of-arrival/age-at-migration fixed effect.[55]

The empirical analysis uses the 1970, 1980, and 1990 US Censuses and is restricted to immigrant men who arrived either in 1965–1969 or in 1975–1979. A cohort is defined in terms of country of birth (85 national origin groups) and age at arrival (25–34, 35–44, and 45–54 years old), and is tracked across the Censuses for a 10-year period. The first column of Table 5 reports the estimated θ. There is a *positive*, though insignificant, unconditional correlation between the rate of wage growth and the log entry wage of immigrant cohorts. The point estimate suggests that the earnings of different immigrant groups diverge somewhat over time – the cohorts that have the highest log wage at the time of entry experience a slightly faster rate of wage growth. In other words, there seems to be some weak relative complementarity between the skills that immigrants bring into the United States and the skills that they acquire in the post-migration period. This result, of course, resembles Mincer's (1974) finding of complementarity between investments in school and investments in on-the-job training.

To evaluate the presence of conditional convergence, consider the regression model:

$$\Delta w_{jk}(t, t') = \theta^* w_{jk}(t) + \phi s_{jk}(t) + \xi_{kt} + \omega_{jk}, \tag{68}$$

where $s_{jk}(t)$ gives the average years of schooling of the immigrant cohort that originated

[54] This concept plays an important role in the economic growth literature (Barro, 1991; Barro and Sala-i-Martin, 1992). In this literature, per-capita income across countries converges if the initial level of the human capital stock is held constant across countries, but does not converge if initial human capital varies across countries.

[55] The inclusion of the fixed effect ξ_{kt} in (67) implies that the numerical value of the coefficient θ is unchanged if the dependent variable were redefined to be the rate of wage growth of the immigrant cohort relative to that experienced by natives in the same age group, and the independent variable were the log entry wage of the immigrant cohort minus the log wage of natives in that age group.

Table 5
Convergence regressions in the United States[a]

Independent variable	Dependent variable: rate of wage growth in first 10 years in the United States			
	(1)	(2)	(3)	(4)
Log wage at time of entry	0.049	− 0.428	− 0.711	− 0.824
	(0.121)	(0.074)	(0.067)	(0.065)
Average years of	–	0.050	–	0.045
schooling at time of entry		(0.006)		(0.007)
Fixed effects for country of origin	No	No	Yes	Yes
R^2	0.301	0.648	0.820	0.840

[a] Notes: Standard errors reported in parentheses. The regressions are estimated in the sample of men aged 25–64, who work in the civilian sector, who are not self-employed, and who do not reside in group quarters. The unit of observation is an immigrant cohort, defined in terms of country of origin, age-at-arrival, and calendar year-of-arrival. The cohorts included in the regression arrived either between 1965–1970 or between 1975–1980. All regressions also include a vector of fixed effects indexing a particular age-at-arrival/calendar-year-of-arrival group. The regressions have 414 observations. See Borjas (1999) for details.

from country j at age k – measured as of the time of entry t. The second column of Table 5 shows that θ^*, a measure of conditional convergence, is negative and significant. The same sign reversal occurs if the regression adds country-of-origin fixed effects (see column 3), so that there is a great deal of convergence among immigrant groups from a particular country of origin. These country-of-origin fixed effects, of course, can also be interpreted as measures of the cohort's human capital stock at the time of entry.

Duleep and Regets (1997) have estimated these types of convergence regressions but use a different definition of an immigrant cohort. In particular, the immigrant cohort is defined not only in terms of country-of-origin, age-at-migration, and year-of-arrival (i.e., a cell in j, k, t), but also in terms of educational attainment. In particular, let $w_{jkS}(t)$ be the log wage of an immigrant cohort originating in country j, migrating at age k, with s years of schooling, and arriving in calendar year t. Similarly, let $\Delta w_{jkS}(t, t')$ be the rate of wage growth experienced by this cohort over the time interval (t, t'). For expositional convenience, suppose that all immigrant cohorts arrive in the same calendar year t. Consider the regression model:

$$\Delta w_{jkS} = \lambda w_{jkS} + \xi_k + \omega_{jkS}, \tag{69}$$

where ω_{jkS} is an i.i.d. error term. Duleep and Regets (1997) document that λ is strongly negative in US data, and interpret this finding as implying that the decline in quality across successive immigrant cohorts is not as strong as suggested by the trend in entry wages. A negative λ suggests that more recent cohorts will experience faster wage growth in the

future, and the present value of the age-earnings profile might not differ much across cohorts.

This alternative framework raises the interesting question of whether the coefficient λ estimates the unconditional rate of convergence (θ) or the conditional rate of convergence (θ^*). To see the relationship among these parameters, rewrite the wage level and wage growth for the (j, k, s) cohort as

$$w_{jkS} = w_{jk} + \varphi_S + e_{jkS}, \tag{70}$$

$$\Delta w_{jkS} = \Delta w_{jk} + \chi_s + \varepsilon_{jkS}, \tag{71}$$

where φ_S and χ_s are fixed effects giving the "returns to schooling" for wage levels and wage growth, respectively; and e_{jkS} and ε_{jkS} are i.i.d. random variables that are uncorrelated with the other right-hand side variables in (70) and (71). The convergence regression in (69) can be rewritten as

$$\Delta w_{jk} = \lambda w_{jk} + (\lambda\varphi_s - \chi_s) + \xi_k + \omega', \tag{72}$$

where $\omega' = \omega_{jkS} + \lambda e_{jkS} - \varepsilon_{jkS}$, and an observation is a (j, k, s) cell. Let $p_{jk}(s)$ be the fraction of the population that has s years of schooling in a (j, k) cell, and aggregate across schooling groups within a (j, k) cell.[56] This aggregation yields

$$\Delta w_{jk} = \lambda w_{jk} + \sum_s (\lambda\varphi_s - \chi_s)p_{jk}(s) + \xi_k + \varpi. \tag{73}$$

Eq. (73) shows that the convergence regression that uses schooling groups to define the cohort is equivalent to a regression that aggregates across schooling groups but includes variables that indicate the educational attainment of the cohort. As a result, the coefficient λ estimates the extent of conditional convergence across immigrant cohorts. It is not surprising, therefore, that Duleep and Regets (1997) find a great deal of wage convergence across immigrant cohorts since they are implicitly holding initial skills constant. It is worth stressing, however, that a finding of conditional convergence does *not* suggest that immigrant cohorts with lower entry wages experience faster wage growth in the host country. As Table 5 shows, the choice of a base group is crucial. Overall, immigrant cohorts that start out with higher wages, if anything, tend to have slightly faster wage growth.

5. Immigration and the wage structure

The literature attempting to measure how immigrants affect the employment opportunities of native workers in a host country has grown rapidly in the past decade. However, a number of difficult conceptual and econometric problems plague this literature. As a result, much of the accumulated empirical evidence probably has little to say about a central question in the economics of immigration.

[56] The aggregation uses $p_{jk}(s)$ as weights.

5.1. Spatial correlations

Economic theory suggests that immigration into a *closed* labor market affects the wage structure in that market by raising the wage of complementary workers and lowering the wage of substitutes. Almost all of the empirical studies in this literature define the labor market along a geographic dimension – such as metropolitan areas or states in the United States. If immigrant flows penetrate geographic labor markets in the host country randomly *and* if natives do not respond to these supply shocks, the "spatial correlation" between labor market outcomes in a locality and the extent of immigrant penetration would identify the impact of immigration. Beginning with the early work of Grossman (1982) and Borjas (1983), the typical study regresses a measure of native economic outcomes in the locality (or the change in that outcome) on the relative quantity of immigrants in that locality (or the change in the relative number).[57] The regression coefficient is then interpreted as the "impact" of immigration on the native wage structure.

There are two well-known problems with this approach. First, immigrants may not be randomly distributed across labor markets. The 1990 US Census indicates that immigrants cluster in a very small number of places: 73.8% of immigrants aged 18–64 reside in 6 states (California, New York, Texas, Florida, Illinois, and New Jersey), but only 35.5% of natives live in those states. Similarly, 35.4% of immigrants live in four metropolitan areas (Los Angeles, New York, Chicago, and Miami), but only 12.9% of natives live in those localities. If the areas where immigrants cluster (e.g., California) have done well over some time periods, this would produce a spurious correlation between immigration and area outcomes either in the cross-section or in the time-series. A positive spatial correlation would simply indicate that immigrants choose to reside in areas that are doing relatively well, rather than measure the extent of complementarity between immigrant and native workers.

The second problem with the spatial correlation approach is that natives may respond to the entry of immigrants in a local labor market by moving their labor or capital to other localities until native wages and returns to capital are again equalized across areas. A large immigrant flow arriving in Los Angeles might well result in, say, fewer workers from Mississippi or Michigan moving to California, and a reallocation of capital from those states to California. A comparison of the wage of native workers between California and other states might show little or no difference because the effects of immigration are diffused throughout the national economy, and not because immigration had no economic effects.

In view of these potential problems it is not too surprising that the empirical literature has produced a confusing array of results. The generic regression model used in the spatial correlation literature is of the form:[58]

[57] More recent studies include Altonji and Card (1991), Card (1997), Jaeger (1996), LaLonde and Topel (1991), and Schoeni (1997). De New and Zimmermann (1994) and Pischke and Velling (1997) provide similar studies of the German labor market.

[58] The early studies estimated Eq. (74) in level form, while more recent studies tend to use first-difference measures of labor market outcomes.

$$\Delta y_{js}(t,t') = \beta_t \Delta m_{js}(t,t') + X_{js}(t)\alpha_t + u_{js}(t,t'), \tag{74}$$

where $\Delta y_{js}(t,t')$ is the change in a measure of employment opportunities experienced by natives who live in region j and belong to skill group s between years t and t'; $\Delta m_{js}(t,t')$ is a measure of the immigrant supply shock in that region for that skill group over the (t,t') time interval; X is a vector of standardizing variables; and $u_{js}(t,t')$ is the stochastic error.

Table 6 summarizes the estimated β's from recent studies by Borjas et al. (1997) and Schoeni (1997). The Borjas–Freeman–Katz study uses states as the geographic unit, covers the 1960–1970, 1970–1980, and 1980–1990 periods, and defines the immigrant supply shock $\Delta m_{js}(t,t')$ as the change in the number of immigrants between t and t' relative to the number of natives in cell (j, s) at time t. Borjas, Freeman, and Katz pool across education groups and estimate Eq. (74) by including fixed effects indicating the native group's educational attainment and state of residence. The Schoeni study uses metropolitan areas as the geographic unit, covers the 1970–1980 and 1980–1990 time periods, and defines the immigrant supply shock as the change in the fraction of the total population that is foreign-born. Schoeni estimates Eq. (74) separately by education group, and includes the native group's mean education and age, as well as a measure of the size of the labor market, in the vector X. In both studies, the immigrant supply shock is related to wage and employment changes.

The most striking feature of Table 6 is that each study finds huge differences across coefficients, making it extremely difficult to generalize about the effect of immigration on labor market outcomes. Both studies report that the sign of the coefficient β_t changes erratically over time. In the Borjas–Freeman–Katz analysis, there is a negative correlation between immigration and employment in the 1960s, but the coefficient becomes positive (and numerically larger) in the 1970s, and turns negative and modest in the 1980s. Similarly, Schoeni finds that a three-point increase in the immigrant share of the population (from, say, 7 to 10%) reduced the earnings of men who are high school graduates by 1% in the 1970s, but the same supply shock would have increased the wage of this group by 0.8% had it occurred between 1980 and 1990. Note also that there is a lot of dispersion in the coefficients (within a given time period) when one compares the results for men and women, or if one looks at wage outcomes or employment outcomes.

As noted above, the supply shock to a particular labor market is likely to be endogenous because immigrants choose where to live depending on economic conditions in the locality (this point is discussed in more detail in the next section). Altonji and Card (1991, p. 222) instrument the immigrant supply shock with a second-order polynomial in the fraction of the work force that is foreign-born at the beginning of the period. In the Altonji–Card study (which covers the 1970–1980 period), the OLS estimate of β_t for white men with less than a high school education is -0.36 (with a standard error of 0.41), but the IV estimate is -1.10 (0.64). The Altonji–Card IV estimate of Eq. (74), therefore, seems to suggest that immigrants have a substantial adverse effect on the wages of natives.

The Schoeni study uses the Altonji–Card IV procedure, and also finds that IV leads to very different estimates. As Table 6 shows, however, the IV procedure does not reduce the

Table 6
Summary of results from spatial correlations approach[a]

Dependent variable/group	Men			Women		
	1960–1970	1970–1980	1980–1990	1960–1970	1970–1980	1980–1990
State data, OLS						
Log weekly earnings	0.59	0.07	−0.10	0.20	0.37	−0.02
	(0.11)	(0.08)	(0.06)	(0.21)	(0.14)	(0.04)
Employment probability	−0.06	0.08	−0.03	0.19	0.11	0.01
	(0.03)	(0.05)	(0.01)	(0.05)	(0.09)	(0.01)
Metropolitan area data, OLS						
Years of schooling < 12						
Log weekly earnings	–	−0.09	0.69	–	−0.77	0.73
	–	(0.29)	(0.32)	–	(0.40)	(0.26)
Labor force participation rate	–	−0.02	−0.21	–	0.13	−0.42
	–	(0.10)	(0.10)	–	(0.15)	(0.10)
Years of schooling = 12						
Log weekly earnings	–	−0.32	0.27	–	0.13	0.86
	–	(0.23)	(0.28)	–	(0.25)	(0.23)
Labor force participation rate	–	0.01	−0.12	–	0.27	−0.21
	–	(0.08)	(0.06)	–	(0.11)	(0.07)
Years of schooling > 12						
Log weekly earnings	–	0.03	0.45	–	0.04	0.83
	–	(0.25)	(0.15)	–	(0.29)	(0.17)
Labor force participation rate	–	−0.08	−0.05	–	0.30	−0.22
	–	(0.07)	(0.04)	–	(0.14)	(0.07)

Metropolitan area data, IV						
Years of schooling < 12						
Log weekly earnings	–	−1.05	1.12	–	−2.72	1.20
	–	(0.42)	(0.36)	–	(0.63)	(0.31)
Labor force participation rate	–	−0.37	−0.23	–	−0.27	−0.43
	–	(0.16)	(0.11)	–	(0.21)	(0.12)
Years of schooling = 12						
Log weekly earnings	–	−0.96	1.01	–	−0.55	1.20
	–	(0.31)	(0.35)	–	(0.35)	(0.27)
Labor force participation rate	–	0.08	−0.20	–	0.50	−0.25
	–	(0.09)	(0.07)	–	(0.15)	(0.08)
Years of schooling > 12						
Log weekly earnings	–	−0.76	0.72	–	−0.39	1.05
	–	(0.29)	(0.18)	–	(0.38)	(0.20)
Labor force participation rate	–	−0.11	0.00	–	0.17	−0.26
	–	(0.11)	(0.10)	–	(0.17)	(0.08)

[a] Notes: Standard errors are reported in parentheses. The regression coefficients from the state data are drawn from Borjas et al. (1997, Table 7), and the regression coefficients from the metropolitan area data are drawn from Schoeni (1997, Tables 1, 2, 3). The IV procedure instruments the immigrant supply shock with a second-order polynomial in the fraction of the work force that is foreign-born at the beginning of the period.

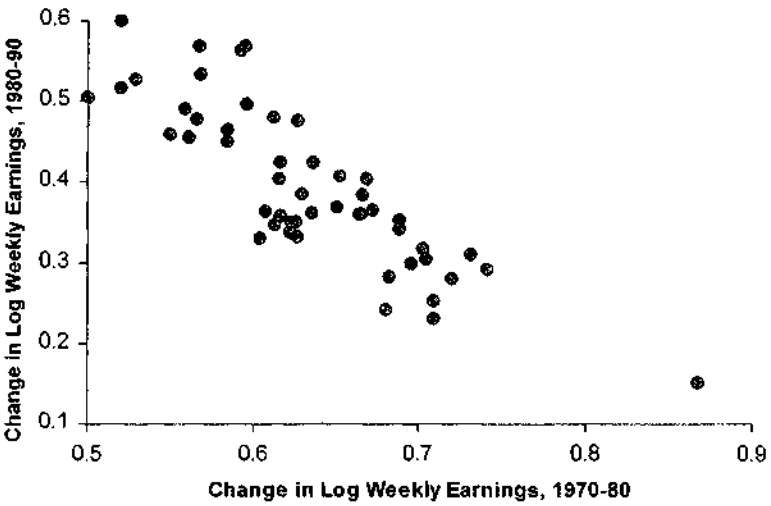

Fig. 2. Wage growth by state, 1980–1990 and 1970–1980.

confusion created by the excessive time variation in the estimated β's. If anything, the IV procedure increases it. In the 1970s, the OLS spatial correlation is usually negative and the IV procedure tends to make β even more negative. In the 1980s, the OLS spatial correlation is usually positive and the IV procedure tends to make β even more positive.

The ambiguous empirical evidence raises a number of important questions – most of which have yet to be seriously addressed by the literature. For instance, why is the sign of the spatial correlation in the United States so dependent on the time period under analysis? Borjas, Freeman, and Katz suggest that the instability in the spatial correlation over time can probably be traced back to major changes in the US regional wage structure – changes that are not well understood and that probably have little, if anything, to do with immigration. Fig. 2 illustrates the nature of the structural change by showing the relationship by state between (education-adjusted) wage growth in the 1980s and wage growth in the 1970s for men.[59] The figure illustrates a strong *negative* correlation in wage growth by state across the two decades.[60] In other words, the high wage growth states of the 1970s became low wage growth states in the 1980s.

However, Fig. 3 shows that *the same states* continued to receive large numbers of immigrants. The reversal of wage growth among states thus implies a reversal in the

[59] The data underlying the figure adjusts for interstate differences in the educational attainment of natives by aggregating across different education cells using a fixed weight of the native education distribution; see Borjas et al. (1997) for more details.

[60] Borjas–Freeman–Katz show that this negative correlation does not exist between the 1960s and the 1970s. The correlation in those two decades is nearly zero. Shoeni (1997, unpublished tabulations) also finds a strong negative correlation in wage growth by metropolitan area between the 1970s and the 1980s.

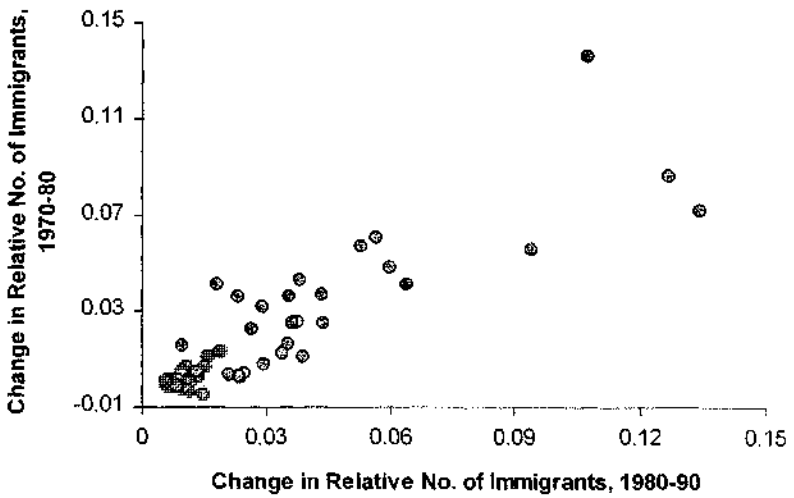

Fig. 3. Immigrant supply shocks by state, 1980–1990 and 1970–1980.

sign of the correlation between changes in wages and in immigration. An observer will almost certainly draw different inferences about the impact of immigration by analyzing spatial correlations estimated in different time periods. Unless the analyst can net out the impact of these structural shifts (and that would require an understanding of why the shifts occurred in the first place), it is almost hopeless to isolate the impact of immigration on the US wage structure from regression-based spatial correlations.

A different approach to estimating spatial correlations appears in Card's (1990) influential case study of the Mariel immigrant flow. On April 20, 1980, Fidel Castro declared that Cuban nationals wishing to move to the United States could leave freely from the port of Mariel. By September 1980, about 125,000 Cubans had chosen to undertake the journey. Almost overnight, the Mariel "natural experiment" increased Miami's labor force by 7%. Card's (1990) analysis of the CPS data indicates that labor market trends in Miami between 1980 and 1985 – in terms of wage levels and unemployment rates – were similar to those experienced by such cities as Los Angeles, Houston and Atlanta, cities that did not experience the Mariel supply shock.[61]

Although superficially different, all spatial correlation studies – whether they use the regression model in (74) or focus on a single unexpected supply shock – rely on difference-in-differences estimates of how immigration changes native outcomes in cities that

[61] Related studies include Hunt's (1992) analysis of the movement of 900,000 persons of European origin between Algeria and France in 1962, and Carrington and de Lima's (1996) study of the 600,000 refugees who entered Portugal after the country lost the African colonies of Mozambique and Angola in the mid-1970s. Neither study finds a substantial impact of immigration on the affected local labor markets.

received immigrants versus in cities that did not.[62] One could easily argue that this literature has failed to increase our understanding of how labor markets respond to immigration. If we take the empirical evidence summarized in Table 6 at face value, the implications are disturbing: either we need different economic models to understand how supply shocks affect labor markets in different time periods (and we would then be left wondering which model we should use to predict the impact of the next immigrant wave), or the regression coefficients are simply not measuring what we think they should be measuring.

5.2. *A model of wage determination and internal migration*

As noted earlier, natives might respond to immigration by "voting with their feet," either through capital or labor flows. What structural parameters, if any, do the spatial correlations between native wages and immigrant supply shocks then measure? And, in particular, is there a way of recovering the "true" wage effect of immigration from spatial correlations?

This section shows formally what these spatial correlations identify in a simple framework that jointly models the wage determination process in a local labor market and the internal migration decision of native workers. The model presented here borrows liberally from a framework developed by Borjas et al. (1997, unpublished appendix).[63]

Suppose that the labor demand function in geographic area j ($j = 1, \ldots, J$) at time t can be written as

$$w_{jt} = X_{jt} L_{jt}^{\eta}, \tag{75}$$

where w_{jt} is the wage in region j at time t; X_{jt} is a demand shifter; L_{jt} gives the total number of workers (both immigrants, M_{jt}, and natives, N_{jt}); and η is the factor price elasticity ($\eta < 0$). It is useful to interpret Eq. (75) as the marginal productivity condition for a group of workers with a particular skill level. For convenience, I omit the subscript indicating the skill class, and I assume that all workers within a particular skill class are perfect substitutes.

Suppose that $N_{j,-1}$ native workers reside in region j in the pre-immigration regime ($t = -1$), and that the national labor market is in equilibrium prior to the entry of immigrants. The wage, therefore, is initially constant across all J regions. We can then write the marginal productivity condition in the pre-immigration regime as

$$w_{j,-1} = X_{j,-1} N_{j,-1}^{\eta} = w_{-1}, \qquad \forall j. \tag{76}$$

[62] The key distinction between the two approaches concerns the extent to which the immigrant flow is unexpected (and natives have had little opportunity to plan in advance for the supply shock).

[63] The model can be viewed as an application of the Blanchard and Katz (1992) framework that analyzes how local labor markets respond to demand shocks. The model can also be adapted to incorporate capital flows.

We will assume that this economy is affected only by supply shocks, so that the demand shifter X_{jt} remains constant across all time periods (i.e., $X_{jt} = X_{j,-1}$, $\forall j$).[64]

It is instructive to begin with a very simple version of the supply shock, a one-time supply increase. In particular, M_{j0} immigrants enter region j at time 0. This supply shock will generally induce a response by native workers, but this response occurs *with a lag*. For simplicity, assume that immigrants do not migrate internally within the United States – they enter region j, and remain there.[65] Natives do respond, and region j experiences a net migration of ΔN_{j1} natives in period 1, ΔN_{j2} natives in period 2, and so on. The variable N_{jt} then gives the number of native workers present in region j at time t, and M_{jt} gives the number of immigrants who entered (and remained) in region j. The wage in region j at time t is given by

$$\log w_{jt} = \log X_{jt} + \eta \log(N_{j,-1} + M_{j0} + \Delta N_{j1} + \cdots + \Delta N_{jt}), \tag{77}$$

which can be rewritten as

$$\log w_{jt} \approx \log w_{-1} + \eta(m_{j0} + v_{j1} + \cdots + v_{jt}), \qquad \text{for } t \geq 0, \tag{78}$$

where $m_{j0} = M_{j0}/N_{j0}$, the relative number of immigrants entering region j; and $v_{jt} = \Delta N_{jt}/N_{j0}$, the net migration rate of natives in region j at time t (relative to the initial population in the region).[66]

The lagged native supply response is described by the function

$$v_{jt} = \sigma(\log w_{j,t-1} - \log \bar{w}), \tag{79}$$

where $\log \bar{w}$ is the equilibrium wage that the national economy will attain once the one-time immigrant supply shock works itself through the system, and σ is the supply elasticity ($\sigma > 0$).[67] The equilibrium wage that will be eventually attained in the national economy is defined by:

$$\log \bar{w} = \log w_{-1} + \eta m, \tag{80}$$

[64] This assumption implies that the entry of immigrants will necessarily lower the average wage in the economy. The model can be extended to allow for capital flows from abroad. These capital flows would bring the rental rate of capital back to the world price and re-equilibrate the economy at the pre-migration wage. This extension, however, complicates the notation substantially without altering the key insights.

[65] Some of the "movers" will be immigrants taking advantage of better opportunities in other regions. The empirical evidence in Bartel (1989), however, suggests that immigrants in the United States are not very mobile once they enter the main gateway areas. The possibility that some of the movers might be immigrants does not affect the nature of the results reported below.

[66] The lag in native migration decisions implies that $N_{j0} = N_{j,-1}$.

[67] The supply function is typically written in terms of wage differentials among regions. Consider a two-region framework with equally sized regions. The alternative specification of the supply function is: $v_2 = \gamma(\log w_2 - \log w_1)$, where γ would be the conventionally defined supply elasticity. Because the regions are equally sized, the equilibrium wage $\log \bar{w} = 0.5(\log w_2 + \log w_1)$. Substituting this definition into the supply function yields: $v_2 = 2\gamma(\log w_2 - \log \bar{w})$, so that the elasticity σ defined in (79) is twice the conventionally defined supply elasticity.

where $m = M/N$; M gives the total number of immigrants in the economy; and N gives the (fixed) total number of natives.

The relationship between the region-specific supply shock m_{j0} and the national supply shock, m, is easy to derive. In particular, suppose region j has (in the pre-immigration regime) a fraction r_j of the native population and receives a fraction ρ_j of the immigrants. The region-specific supply shock is then given by

$$m_{j0} = \frac{M_{j0}}{N_{j0}} = \frac{\rho_j M}{r_j N} = k_j m, \tag{81}$$

where $k_j = \rho_j / r_j$, a measure of the penetration of immigrants into region j *relative* to the region's pre-immigration size. Immigration is "neutrally" distributed across the host country if $k_j = 1$, $\forall j$. The long-run equilibrium wage log $\bar{w}$ defined in Eq. (80) would be attained immediately in all regions if the immigrant supply shock were neutrally distributed over the country.

There are a number of substantive assumptions implicit in the supply function given by Eq. (79) that are worth noting. First, the native supply response is lagged. Immigrants arrive in period 0. The demand function in Eq. (78) implies that the wage response to immigration is immediate, so that wages fall in the affected regions. Natives, however, do not respond to this change in the regional wage structure until period 1. Secondly, the model has not imposed any restrictions on the value of the parameter σ. If σ is sufficiently "small," the migration response of natives may not be completed within one period. Some individuals may respond immediately, but other individuals will take somewhat longer.[68] Finally, note that the migration decision is made by comparing the current wage in region j to the wage that region j will eventually attain. In this model, therefore, there is perfect information about the eventual outcome that results from the immigrant supply shock. Unlike the typical cobweb model, persons are not making decisions based on erroneous information. The lags arise simply because it is difficult to change locations immediately.

The model is now closed and can be solved recursively. The native net migration rate in region j at time t is given by[69]

$$v_{jt} = -\eta\sigma(1 + \eta\sigma)^{t-1}(1 - k_j)m, \tag{82}$$

where the restriction $0 < (1 + \eta\sigma) < 1$ is assumed to hold throughout the analysis. Eq.

[68] In a sense, the migration behavior underlying Eq. (79) is analogous to the firm's behavior in the presence of adjustment costs (Hamermesh, 1993). One can justify this staggered response in a number of ways. The labor market is in continual flux, with persons entering and leaving the market, and some of the migration responses may occur concurrently with these transitions. Workers may also face constraints that prevent them from taking immediate advantage of regional wage differentials. Some families, for example, might have children enrolled in school or might lack the capital required to fund the migration.

[69] Eq. (82) is derived as follows. First, use the demand function in (78) to calculate the wage observed in region j at time 0 after the immigrant supply shock. This wage can then be used to calculate the net migration flow experienced by region j in period 1 using the supply function in (79), and to calculate the period-1 wage in the region. Eq. (82) follows from this procedure by carrying the process forward to period t.

(82) shows that region j does not experience any net migration of natives if $k_j = 1$, since the "right" share of immigrants entered that region in the first place. Regions that received a relatively large number of immigrants ($k_j > 1$) experience native out-migration in the post-immigration period (recall $\eta < 0$), while regions that received relatively few immigrants experience native in-migration. Native net migration is largest immediately after the immigrant supply shock, and declines exponentially thereafter.

The wage in region j at time t depends on the total net migration of natives up to that time. This total migration is given by

$$V_{jt} = -\sum_{\tau=1}^{t} \eta\sigma(1+\eta\sigma)^{\tau-1}(1-k_j)m = (1-k_j)[1-(1+\eta\sigma)^t]m. \tag{83}$$

Eq. (78) then implies that the wage in region j at time t equals

$$\log w_{jt} = \log w_{-1} + \eta\left\{k_j + (1-k_j)[1-(1+\eta\sigma)^t]\right\}m. \tag{84}$$

Eqs. (83) and (84) provide the foundations for a two-equation model that jointly analyzes the native response to immigration and the immigrant impact on the wage structure. To evaluate if the data can identify the relevant parameters, consider a slightly different form of the model:

$$V_{jt} = [1-(1+\eta\sigma)^t]m - [1-(1+\eta\sigma)^t]m_j, \tag{85}$$

$$\log w_{jt} - \log w_{-1} = \eta[1-(1+\eta\sigma)^t]m + \eta(1+\eta\sigma)^t m_j. \tag{86}$$

Note that both Eqs. (85) and (86) are of the "before-and-after" type. In effect, Eq. (85) presents a first-difference model of the total migration of natives (where there was zero migration in the pre-immigration regime), while Eq. (86) presents a model of the wage change in region j before and after the immigrant supply shock. Both regressions contain two explanatory variables: the national immigrant supply shock (m), and the regional supply shock (m_j). The model has been derived for a single skill class, so that the national immigrant supply shock is a constant across all observations and its coefficient is subsumed into the intercept. One can imagine having a number of different skill classes and "stacking" the data across skill groups (assuming that there are no cross-effects that must be taken into account). The national immigrant supply variable would then be a constant within a skill class. It is likely, however, that there are skill-specific fixed effects both in net migration rates and in wage changes. These fixed effects imply that the coefficient of the national supply shock cannot be separately estimated. Therefore, all the estimable information about how regional wages evolve and how natives respond to immigration is contained in the coefficient of the supply shock variable m_j.

Suppose we observe data as of time t (i.e., t years after the immigrant supply shock). Let δ_t be the coefficient from the native net migration regression, and β_t be the coefficient from the wage change regression. These coefficients are defined by

$$\delta_t = -[1 - (1 + \eta\sigma)^t], \tag{87}$$

$$\beta_t = \eta(1 + \eta\sigma)^t. \tag{88}$$

These coefficients yield a number of interesting implications. As t grows large, the coefficient in the migration regression converges to -1 and the coefficient in the wage change regression converges to zero. Put differently, the longer the time elapsed between the one-time immigrant supply shock and the measurement of native migration decisions and wage changes, the more likely that natives have completely internalized the supply shock, and the less likely that the data will uncover *any* wage effect on local labor markets. Second, note that the wage regression will not estimate the factor price elasticity η except at time 0 – *immediately* after the immigrant supply shock. Over time, the wage effect is contaminated by native migration, and the contamination grows larger the longer one waits to measure the effect. In fact, reasonable assumptions for the factor price and supply elasticities suggest that the wage regression will yield useless estimates of the wage effect even if the data is observed only 10 years after the one-time supply shock. For example, suppose that $\eta = -0.3$, and that $\sigma = 0.5$. After 10 years, the wage change regression would yield a coefficient of -0.06. Finally, and most important, the two-equation model allows us to identify the factor price elasticity if we do not wait "too long" after the immigrant supply shock. The definitions of the coefficients δ_t and β_t imply that

$$\eta = \frac{\beta_t}{1 + \delta_t}. \tag{89}$$

The factor price elasticity can be estimated from the spatial correlation between wage growth and immigration by "blowing up" the coefficient from the wage change regression. Suppose, for example, that the migration coefficient is -0.5, so that 5 natives leave the region for every 10 "excess" immigrants that enter. The true factor price elasticity η is then estimated by doubling the spatial correlation between wages and immigration. Note, however, that because δ approaches -1 as t grows large, the formula given by Eq. (89) is not useful if the data are observed some time after the immigrant supply shock took place.[70]

The model suggests that the problem with the spatial correlations reported in the literature may not be so much the endogeneity problem caused by immigrants choosing to move to "good" areas, but the fact that all of the currently available empirical models suffer from omitted-variable bias. The correct specification of the wage change regression is one in which the wage change in the region (for a particular skill group) is regressed on the *net* supply shock induced by immigration. The correct generic regression is of the form

$$\Delta w_j = \eta(m_j + V_j) + \text{other variables} + e_j, \tag{90}$$

where m_j measures the immigrant supply shock; V_j measures the (total) net migration rate of natives; and e_j is the stochastic error. The typical regression in the literature is of the form

$$\Delta w_j = \beta m_j + \text{other variables} + (e_j + \eta V_j). \tag{91}$$

As discussed above, it is not uncommon to estimate Eq. (91) using instrumental variables, where the instrument is the fraction of region j's population that is foreign-born at the beginning of the period. The joint model of wage determination and internal migration, however, clearly indicates that this instrument is invalid because it *must be* correlated with the disturbance term in (91). After all, the native net migration response depends on the number of immigrants in the local labor market at the beginning of the period. As a result, the IV methodology commonly used in the literature does not identify any parameter of interest. A valid IV procedure would require constructing an instrument that is correlated with the immigrant supply shock, but is uncorrelated with the native migration response. Such an instrument, it is fair to say, will be hard to find.[71]

The model also suggests that the factor price elasticity *is* directly identifiable from a before-and-after wage change regression if the regression is estimated immediately after the immigrant supply shock takes place. Card's (1990) study of the Mariel flow carries out precisely this type of exercise, yet fails to find any measurable response to immigration in the Miami labor market in the year after the supply shock took place. Card also reports

[70] Although the model presented here focuses on the response of native workers to immigration, the framework can be extended to take into account the response of capital flows. These capital flows would include both the response of native-owned capital "residing" in other regions, as well as the response of international capital to the lower wages now available in the host country. It is instructive to sketch a model that incorporates these capital flows, and to compare the key results to those of the internal migration model. Let F_{jt} be the capital flow in year t induced by the immigrant supply shock in year 0, and suppose that the supply response of capital is given by: $F_{jt} = \alpha_1(\log w_{jt} - \log \bar{w}_t) + \alpha_2(\log w_{jt} - \log w_{-1})$, where $\bar{w}_t$ gives the average wage observed in the host country at time t. The first term of this equation summarizes the incentives for capital flows to occur within the host country, while the second term summarizes the incentives for international capital flows (assuming that the world economy was in equilibrium at wage w_{-1} prior to the immigrant supply shock.). Note that both supply elasticities α_1 and α_2 are negative. The specification of the capital supply response implies that internal and international capital flows continue until the wage in all regions of the host country re-equilibrate at the world wage w_{-1}. The variable F_{jt} enters additively into the earnings function in (78). To simplify, suppose that there are only capital responses to immigration (and no native internal migration). After some tedious algebra, it can be shown that the equation giving the change in the log wage between time t and -1 (the before-and-after comparison) depends on both m, the national supply shock, and on m_j, the regional supply shock. The coefficient of the regional supply shock (the only coefficient that can be identified by the data) is then given by $\eta(1 + \alpha_1 + \alpha_2)^t$. As with the native migration model, therefore, the factor price elasticity is identifiable only in the initial year, and the spatial correlation converges to zero (assuming that $-1 < \alpha_1 + \alpha_2 < 0$). This approach can be extended to incorporate both native internal migration and capital flows into the model. The simple form of the "blowing up" property reported in Eq. (89) does not hold in this more general model because the true factor price elasticity cannot be identified from estimates of the spatial correlations (β) and the native migration response (δ). The identification of η now also requires information on the elasticities of the capital supply equation.

[71] The generic model in Eq. (90) can be used to illustrate that the "blowing-up" result is a general property of this type of framework. In addition to the wage change equation in (90), there exists an equation relating the native response to the initial supply shock: $V_j = \delta m_j + \text{other variables} + \nu_j$. Substituting this equation into (90) yields the reduced-form regression: $\Delta w_j = \eta(1 + \delta)m_j + \text{other variables} + \omega_j$. The coefficient of m_j in this reduced-form equation equals β, the spatial correlation typically reported in the literature. It then follows that $\eta = \beta/(1 + \delta)$.

evidence that population flows into the Miami area slowed down as a result of the Mariel shock, but it seems unlikely that native migration decisions completely internalized the impact of the supply shock within a year. It is possible that capital flows from other cities to Miami "take up the slack," but there does not exist any evidence indicating that this, in fact, happened. Card's evidence (although imprecisely estimated), therefore, cannot be easily dismissed and the findings of the Card study remain a major puzzle.

5.3. *A model with a permanent supply shock*

The model presented in the previous section assumed that immigration is a one-time supply shock, and the model's parameters were estimated by comparing outcomes in the pre- and post-immigration periods. Some host countries, particularly the United States, have been receiving a continuous (and large) flow of immigrants for more than 30 years. As a result, it is useful to determine what, if anything, can be learned from spatial correlations when immigrants add to the labor supply of the host country in every period, and the parameters of the model are estimated while the immigrant supply shock continues to take place.

The framework presented in the previous section can be easily generalized to the case of a permanent influx if we assume that each region of the country receives the *same* immigrant supply shock every year. This assumption is not grossly contradicted by the data for the United States because the same regions have been the recipients of immigrants for several decades. At time t, therefore, native workers respond to the supply shock that occurred in the preceding period, as well as to the supply shocks that occurred in all earlier periods. The main adjustment that has to be made to the earlier model concerns the specification of the native supply function. In particular, suppose that the native migration response at time t is

$$v_{jt} = \sigma(\log w_{j,t-1} - \log \bar{w}_{t-1}), \tag{92}$$

where $\log \bar{w}_{t-1}$ is the equilibrium wage that will be observed throughout the national economy once all the immigrant supply shocks that have occurred up to time $t-1$ work themselves through the system. As before, the native response is forward-looking in the sense that natives take into account the consequences of the *total* immigrant supply shock that has already taken place. It might seem preferable to model the supply function so that natives take into account the expected impact of future immigration. However, the total supply shock up to time $t-1$ is a "sufficient statistic" because we have assumed that the region receives the same number of immigrants in every period.

The national equilibrium wage that will be eventually attained as a result of the immigrant supply shocks up to period $t-1$ is

$$\log \bar{w}_{t-1} = \log w_{-1} + \eta(m_{j0} + \cdots + m_{j,t-1}) = \log w_{-1} + \eta t m_j. \tag{93}$$

Consider the native supply response to the immigrants who entered the country in period 0. Eq. (83) in the previous section showed that the net migration rate of natives

in period t induced by the period 0 immigrant flow equals $(1 - k_j)[1 - (1 + \eta\sigma)^t]m$. Consider now the native response to the supply shock in year 1. Eq. (83) then implies that the net migration rate of natives induced by the period-1 migration flow equals $(1 - k_j)[1 - (1 + \eta\sigma)^{t-1}]m$. The total net migration of natives in period t attributable to a supply shock of $k_j m$ in region j between periods 0 and $t - 1$ is then given by

$$V_{jt} = \sum_{\tau=0}^{t-1} (1 - k_j)[1 - (1 + \eta\sigma)^{\tau}]m = (1 - k_j)\left[t + \frac{1 + \eta\sigma}{\eta\sigma}[1 - (1 + \eta\sigma)^t]\right]m, \tag{94}$$

and the wage observed in region j at time t equals

$$\log w_{jt} = \log w_{-1} + \eta\left\{(t + 1)k_j + (1 - k_j)\left[t + \frac{1 + \eta\sigma}{\eta\sigma}[1 - (1 + \eta\sigma)^t]\right]\right\}m. \tag{95}$$

We can now derive the two first-difference regression models that compare native net migration rates and wages before-and-after the beginning of the immigrant supply shock. These regression models are given by

$$V_{jt} = \left[\frac{t}{t + 1} + \frac{(1 + \eta\sigma)}{\eta\sigma}\frac{[1 - (1 + \eta\sigma)^t]}{(t + 1)}\right](t + 1)m$$
$$- \left[\frac{t}{t + 1} + \frac{(1 + \eta\sigma)}{\eta\sigma}\frac{[1 - (1 + \eta\sigma)^t]}{(t + 1)}\right](t + 1)m_j, \tag{96}$$

$$\log w_{jt} - \log w_{-1} = \eta\left[\frac{t}{t + 1} + \frac{(1 + \eta\sigma)}{\eta\sigma}\frac{[1 - (1 + \eta\sigma)^t]}{(t + 1)}\right](t + 1)m$$
$$+ \eta\left[\frac{1}{t + 1} - \frac{(1 + \eta\sigma)}{\eta\sigma}\frac{[1 - (1 + \eta\sigma)^t]}{(t + 1)}\right](t + 1)m_j, \tag{97}$$

where the independent variables have been defined to measure the total (as of time t) immigrant supply shock either at the national level, $(t + 1)m$, or at the regional level, $(t + 1)m_j$. As before, we can estimate these models either within a single skill group, or by "stacking" across skill groups. If the latter model also includes skill fixed effects, the regression models can only identify the coefficient of $(t + 1)m_j$. If we let δ_t be the coefficient of the regional supply shock in the internal migration regression, and β_t be the coefficient in the wage change regression, we can estimate

$$\delta_t = -\left[\frac{t}{t + 1} + \frac{(1 + \eta\sigma)}{\eta\sigma}\frac{[1 - (1 + \eta\sigma)^t]}{(t + 1)}\right], \tag{98}$$

$$\beta_t = \eta\left[\frac{1}{t + 1} - \frac{(1 + \eta\sigma)}{\eta\sigma}\frac{[1 - (1 + \eta\sigma)^t]}{(t + 1)}\right]. \tag{99}$$

Eqs. (98) and (99) indicate that the permanent supply shock model yields insights similar to those obtained in the one-time model. In particular, the wage change regression will estimate the factor price elasticity η only at the very beginning of the immigrant supply shock (when $t = 0$). As t grows larger, the coefficient in the migration regression converges to -1, while that of the wage change regression converges to zero. Finally, the manipulation of Eqs. (98) and (99) reveals that $\eta = \beta_t/(1 + \delta_t)$, so that we can still recover the true factor price elasticity from the spatial correlation by blowing up the estimated wage effect – as long as we do not wait too long into the immigration period.

Few empirical studies actually conduct the "before-and-after" regression analysis suggested by Eqs. (98) and (99). The historical data are usually hard to obtain, particularly if the immigrant supply shock has been in motion for some decades. Instead, most empirical studies attempt to estimate the parameters of interest by first-differencing the data, so that all the observations come from the post-migration period. The first-difference models are given by

$$V_{jt} - V_{j,t-1} = [1 - (1 + \eta\sigma)^t]m - [1 - (1 + \eta\sigma)^t]m_j, \tag{100}$$

$$\log w_{jt} - \log w_{j,t-1} = \eta[1 - (1 + \eta\sigma)^t]m + \eta(1 + \eta\sigma)^t m_j, \tag{101}$$

where the independent variables are defined to be the per-period immigrant supply shock.

As before, let $\delta_t = -[1 - (1 + \eta\sigma)^{t-1}]$, the coefficient of m_j in the first-difference native migration equation; and $\beta_t = \eta(1 + \eta\sigma)^t$, the respective coefficient in the first-difference wage equation.[72] Both of these coefficients are negative so that first-difference regressions should have the "right" sign even when all of the data are observed while the immigrant supply shock is under way. Neither of these coefficients, however, estimates a parameter of interest. Moreover, δ_t approaches minus one and β_t approaches zero as $t \to \infty$. As a result, some local labor markets could be the recipients of very large and permanent supply shocks, but spatial correlations will not reveal the impact of these flows on the wage structure if the first-difference regression is estimated some time after the immigrant supply shock began. Finally, the definitions of δ_t and β_t indicate that the factor price elasticity is estimated by blowing up the coefficient from the wage regression, so that $\eta = \beta_t/(1 + \delta_t)$.

5.4. Immigration and native internal migration

The empirical studies that measure spatial correlations typically ignore the fact that identification of the labor market effects of immigration requires the joint analysis of labor market outcomes and the native response to the immigrant supply shock. The few studies that specifically attempt to determine if native migration decisions are correlated with immigration have yielded a confusing set of results. Filer (1992) finds that metropo-

[72] Interestingly, these coefficients are similar to those obtained in the before-and-after regression in the one-period supply shock model (see Eqs. (87) and (88)).

litan areas where immigrants cluster had lower rates of native in-migration and higher rates of native out-migration in the 1970s, and Frey (1995) and Frey and Liaw (1996) find a strong negative correlation between immigration and the net migration rates of natives in the 1990 Census. In contrast, White and Zai (1993) and Wright et al. (1997) report a positive correlation between the in-migration rates of natives to particular cities and immigration flows in the 1980s.

Recent work by Borjas et al. (1997) and Card (1997) provide the first attempts to jointly analyze labor market outcomes and native migration decisions. In view of the disagreement in earlier research, it should not be too surprising that these two studies reach very different conclusions. Card reports a slight positive correlation between the 1985–1990 rate of growth in native population and the immigrant supply shock by metropolitan area. while Borjas et al. (1997) report a strong negative correlation between native net migration in 1970–1990 and immigration by states. The two studies provide a stark example of how different conceptual approaches to the question can lead to very different answers.

Perhaps the clearest evidence of a *potential* relation between immigration and native migration decisions in the United States is summarized in Table 7.[73] Divide the country into three "regions": California, the other five states that receive large numbers of immigrants (New York, Texas, Florida, New Jersey, and Illinois), and the remainder of the country. Table 7 reports the proportion of the total population, of natives, and of immigrants living in these areas from 1950 to 1990. The modern-era immigrant supply shock in the United States began around 1970 and has continued since. It seems natural to contrast pre-1970 changes in the residential location of the native population with post-1970 changes to assess the effects of immigration on native location decisions.

The data reveal that the share of natives who lived in the major immigrant receiving state, California, was rising rapidly prior to 1970. Since 1970, however, the share of natives living in California has barely changed. However, California's share of the *total* population kept rising from 10.2% in 1970 to 12.4% in 1990. Put differently, an extrapolation of the demographic trends that existed before 1970 – *before the immigrant supply shock* – would have predicted the state's 1990 share of the total population quite well.[74] This result resembles Card (1990, p. 255) conclusion about the long-run impact of the Mariel flow on Miami's population. Card estimates that Miami's population grew at an annual rate of 2.5% in the 1970s, as compared to a growth rate of 3.9% for the rest of Florida. After the Mariel low, Miami's annual growth rate slowed to 1.4%, as compared to 3.4% in the rest of Florida. As a result, the actual population of Dade county in 1986 was roughly the same as the pre-Mariel projection made by the University of Florida.

The finding that the rate of total population growth in areas affected by immigrant supply shocks seems to be independent of immigration may have profound implications for the interpretation of spatial correlations between native economic outcomes and immi

[73] This section is based on the discussion by Borjas et al. (1997).

[74] Borjas et al. (1997, Fig. 4) show that the data point for California (and, in fact, for all the other major immigrant-receiving states) lies close to the regression line linking the 1970–1990 population growth rate to the 1950–1970 rate.

Table 7
Regional distribution of adult-age US population, 1950–1990[a]

	California	Other immigrant states	Rest of country
Percent of total US population			
1950	7.2	26.9	65.9
1960	8.9	27.3	63.7
1970	10.2	27.1	62.7
1980	10.9	26.7	62.4
1990	12.4	27.0	60.7
Percent of native US population			
1950	6.9	25.4	67.7
1960	8.6	26.2	65.2
1970	9.6	26.2	64.2
1980	9.7	25.6	64.8
1990	10.0	25.5	64.4
Percent of foreign-born US population			
1950	10.4	44.4	45.2
1960	14.6	44.9	40.6
1970	20.1	43.8	36.0
1980	27.2	41.9	30.9
1990	33.8	40.0	26.1

[a] Source: Borjas et al. (1997, Table 8). The calculations use the 1950–1990 US Censuses. The adult-age population contains all persons aged 18–64 who are not living in group quarters.

gration. In particular, the immigrants who chose a particular area as their destination "displaced" the native net migration that would have occurred, and this native feedback effect diffused the economic impact of immigration from that area to the rest of the country.

To determine the formal relationship between native migration and immigration, define

$$\Delta n_j(t, t') = \frac{N_j(t') - N_j(t)}{L_j(t)} \div (t' - t), \tag{102}$$

$$\Delta m_j(t, t') = \frac{M_j(t') - M_j(t)}{L_j(t)} \div (t' - t), \tag{103}$$

where $N_j(t)$ gives the number of natives living in area j at time t; $M_j(t)$ gives the number of immigrants; and $L_j(t) = N_j(t) + M_j(t)$. The variable $\Delta n_j(t, t')$ gives the (annualized) rate of native population growth in area j between years t and t' relative to the initial population of

Table 8
Regression coefficients estimating the response of change in native population to immigrant supply shocks in the United States, by state[a]

First-difference regression, 1970–1990	Double-difference regressions	
	1970–1990 relative to 1960–1970	1970–1990 relative to 1950–1970
0.777	−0.756	−1.673
(0.311)	(0.278)	(0.285)

[a] Source: Borjas et al. (1997, Table 8). Standard errors reported in parentheses. The regressions have 51 observations (one for each state plus the District of Columbia), except for the regression in the last column, which omits Alaska and Hawaii and has 49 observations.

the area; and $\Delta m_j(t,t')$ gives the annualized contribution of immigrants to population in the area, again relative to the initial population in the area. Card (1997) and Borjas et al. (1997) suggest the regression model:

$$\Delta n_j(t,t') = a + \delta^* \delta m_j(t,t') + e_j. \tag{104}$$

The coefficient δ^* measures the impact of an additional immigrant arriving in region j in the time interval (t,t') on the change in the number of natives living in that region. The coefficient δ^*, therefore, is the empirical counterpart of the parameter δ in the model presented in the previous sections.

Table 8 reports the estimates of Eq. (104) using US states as the geographic unit. The table summarizes the substantive content of the evidence reported in the Borjas–Freeman–Katz (from which Table 8 is drawn) as well as, to some extent, in the Card study. The first column reports that the coefficient δ^* is positive and significant over the 1970–1990 period. This positive correlation between immigration and native net migration is also reported in the Card study, which uses a different empirical specification: the period under analysis is 1985–1990, the geographic region is the metropolitan area, and the analysis distinguishes among skill groups. Despite the differences between the two studies, the conclusion is similar – the same areas tend to attract both immigrants and natives.

The positive correlation seems to imply that natives do not respond to immigration or that perhaps natives even respond by moving *to* areas penetrated by immigrants. Borjas, Freeman, and Katz argue that the regression specification in Eq. (104) misses an important part of the story. In particular, it compares native population growth among states with different levels of immigration between 1970 and 1990, rather than native population growth in a state *before and after* the immigrant supply shock. In other words, the regression model implicitly assumes that each state would have had the same rate of native

population growth in the absence of immigration. But if each state had its own growth path prior to immigration and that growth path *would have continued* absent immigration, the regression might give a misleading inference about immigration's effects. Borjas, Freeman, and Katz thus propose the "double-difference" model:

$$\Delta n_j(t, t') - \Delta n_j(t_0, t_1) = \alpha + \tilde{\delta}[\Delta m_j(t, t') - \Delta m_j(t_0, t_1)] + v_j, \tag{105}$$

where the time interval (t_0, t_1) occurs in the period prior to the immigrant supply shock, and the coefficient $\tilde{\delta}$ measures the impact of an increase in the number of immigrants on the number of natives – relative to the "pre-existing conditions" in the state.

The second column of Table 8 reports the coefficient from the double-difference model using the state's population growth from 1960 to 1970 to measure the pre-existing trend. The estimated $\tilde{\delta}$ is not significantly different from -1, suggesting considerable displacement. Finally, the third column of the table re-estimates the double-difference model using the state's growth rate between 1950 and 1970 to control for pre-existing conditions. This regression yields an even more negative coefficient. Because the estimated $\tilde{\delta}$ is near (or below) -1, the model presented in the previous sections implies that it is impossible to blow up the spatial correlations and calculate the "true" factor price elasticity.

Table 8 shows that whether one finds a negative or a positive impact of immigration on native net migration depends on the counterfactual posed by a particular regression model. The single-difference regression model in Eq. (104) ignores valuable information provided by the state's demographic trends prior to the immigrant supply shock *and* assumes that all states lie on the same growth path in the post-migration period. The double-difference regression model in Eq. (105) accounts for the pre-existing trends *and* assumes that the trends would have continued in the absence of immigration. The specification of a clear counterfactual is crucial in measuring and understanding the link between immigration, native migration decisions, and the impact of immigrants on the wage structure.

Although the data suggest that the total population growth in a state is independent of immigration, the migration response of natives would completely diffuse the effect of immigration only if the native flows of particular skill groups counterbalanced the immigrant influx and left unchanged the relative factor proportions *within* a state. The evidence on this issue, however, is inconclusive. Borjas et al. (1997, Table 10), for instance, report that factor proportions were converging across states even before the immigrant supply shock began circa 1970. As a result, the sign of the correlation between native migration flows in particular skill groups and the corresponding immigrant supply shock depends not only on whether the counterfactual specifies a before-and-after comparison, but also on whether the model controls for the pre-immigration convergence trends.

Finally, all of the empirical studies in the literature fail to take into account the possibility that the response to immigration includes the movement of capital flows to regions affected by immigrant supply shocks. As a result, the joint analysis of native migration decisions and labor market outcomes may not solve the problems with the spatial correlation approach.

5.5. The factor proportions approach

Because the native response to immigration implies that spatial correlations may not estimate the impact of immigration on the labor market, Borjas et al. (1992) proposed an alternative methodology. The "factor proportions approach" compares a nation's actual supplies of workers in particular skill groups to those it would had had in the absence of immigration, and then uses outside information on the elasticity of substitution among skill groups to compute the relative wage consequences of the supply shock.[75]

Suppose the aggregate technology in the host country can be described by a linear homogeneous CES production function with two inputs, skilled labor (L_S) and unskilled labor (L_u):

$$Q_t = A_t[\alpha L_s^{\rho} + (1 - \alpha)L_u^{\rho}]^{1/\rho}. \tag{106}$$

The elasticity of substitution between skilled and unskilled workers is given by $\sigma = 1/(1 - \rho)$. Suppose further that relative wages are determined by the intersection of an inelastic relative labor supply function with the downward-sloping relative labor demand function derived from the CES. Relative wages in year t are then given by

$$\log(w_{st}/w_{ut}) = D_t - \frac{1}{\sigma}\log(L_{st}/L_{ut}), \tag{107}$$

where D_t is a relative demand shifter.

The aggregate supply of skill group j at time t is composed of native workers (N_{jt}) and immigrant workers (M_{jt}):

$$L_{jt} = N_{jt} + M_{jt} = N_{jt}(1 + m_{jt}), \tag{108}$$

where $m_{jt} = M_{jt}/N_{jt}$. Eq. (107) can be rewritten as

$$\log(w_{st}/w_{ut}) = D_t - \frac{1}{\sigma}\log(N_{st}/N_{ut}) - \frac{1}{\sigma}[\log(1 + m_{st}) - \log(1 + m_{ut})]. \tag{109}$$

An immigrant supply shock in the (t, t') time interval changes the relative number of immigrants by $\Delta\log(1 + m_{jt})$ for skill group j. The predicted impact of the immigrant supply shock on the relative wage of skilled and unskilled workers equals

$$\Delta\log(w_{st}/w_{ut}) = -\frac{1}{\sigma}[\Delta\log(1 + m_{st}) - \Delta\log(1 + m_{ut})]. \tag{110}$$

The calculation implied by (110) requires: (a) the aggregation of heterogeneous workers into two skill groups; (b) the assumption that natives and immigrants within each skill group are perfect substitutes; (c) information on the change in the relative number of immigrants for each skill group; and (d) an estimate of the relative wage elasticity ($-1/\sigma$).

The factor proportions literature often assumes that workers with the same educational

[75] Related applications of the factor proportions approach include Freeman (1977), Johnson (1970), and Welch (1969, 1979).

Table 9
The impact of immigration on the United States using the factor proportions approach[a]

	Definition of skill groups	
	High school dropouts and high school graduates	High school equivalents and college equivalents
Relative number of post-1979 unskilled immigrants in 1995 ($m_{ut} = M_{ut}/N_{ut}$)	0.207	0.056
Relative number of post-1979 skilled immigrants in 1995 ($m_{St} = M_{St}/N_{St}$)	0.041	0.043
Log change in relative supplies $= \log(1 + m_{St}) - \log(1 + m_{ut})$	−0.149	−0.013
Estimate of relative wage elasticity	−0.322	−0.709
Change in log relative wage attributable to post-1979 immigration	0.048	0.009
Actual change in log relative wage between 1980 and 1995	0.109	0.191

[a] Source: Borjas et al. (1997, Tables 14 and 18).

attainment are perfect substitutes.[76] Table 9 summarizes the results from the most recent application of this approach by Borjas et al. (1997), using two alternative classifications of skill groups. In the first, workers who are high school dropouts are defined to be "unskilled," and all other workers are defined to be "skilled." In the second, the skill groups are defined in terms of high school equivalents versus college equivalents. To isolate the labor market effects of post-1979 immigration, the simulation normalizes the data so that all persons present in the United States as of 1979 are considered "natives." The immigrant supply shock that occurred between 1980 and 1995 increased relative supplies by 20.7 percentage points for high school dropouts, and by 4.1 percentage points for workers with at least a high school education. The change in the log gap defined by the bracketed term in (110) is −0.149. Borjas et al. (1992) estimate the relative wage elasticity for these two groups to be −0.322. Eq. (110) then implies that the immigration-induced change in the relative supply of high school dropouts reduced their relative wage by 4.8 percentage points, or about 44% of the total decline in the relative wage of high school dropouts between 1980 and 1995.

Table 9 also shows, however, that immigration has a much smaller impact if we use an alternative skill aggregation. The post-1979 immigrants increased the relative supply of high-school equivalents by only 1.3 percentage points. Katz and Murphy (1992) estimate that the relative wage elasticity for these two groups is −0.709. The immigrant supply

[76] Jaeger (1996) presents evidence that immigrant and native workers within broadly defined education groups may be near-perfect substitutes.

shock then lowered the college/high school wage differential by about 0.9 percentage points, about 5% of the actual decline in this wage gap.

In an important sense, the factor proportions approach is unsatisfactory. It departs from the tradition of decades of research in labor economics that attempts to estimate the impact of a particular shock on the labor market by directly observing how this shock affects some workers and not others. The factor proportions approach does not *estimate* the impact of immigration on the wage structure; rather, it *simulates* the impact. For a given elasticity of substitution, the factor proportions approach mechanically predicts the relative wage consequences of a supply shock. It is not surprising that the approach has been criticized for relying on theoretical models to calculate the effect of immigration on native outcomes (Card, 1997, p. 2; DiNardo, 1997, p. 75).

On the one hand, the criticism is valid. The factor proportions approach certainly relies on a theoretical framework. If the model of the labor market underlying the calculations or the estimate of the relative wage elasticity is incorrect, the estimated impact of immigration is also incorrect. On the other hand, a great deal of empirical research shows that relative supplies *do* affect relative prices.[77] Moreover, the spatial correlations estimated over the past 15 years have failed to reveal with any degree of precision the impact that immigration has on the wage structure. Finally, although the factor proportions approach relies on theory, so must any applied economic analysis that wishes to do more than simply calculate correlations. In the end, *any* interpretation of economic data – and particularly any use of these data to predict the outcomes of shifts in immigration policy – requires a "story". The factor proportions approach tells a very specific story of the economy and relies on that story to estimate the impact of immigration on the wage structure.

6. Conclusion

Our understanding of the labor market effects of immigration grew significantly in the past two decades. In view of the potential policy implications of this research and the emotional questions that immigration raises in many countries, it is inevitable that these advances have been marked by heated and sometimes contentious debate over a number of conceptual and methodological issues. Nevertheless, we now have a better grasp on a number of central questions: Which types of persons choose to emigrate? What is the relative importance of aging and cohort effects in determining how the skills of immigrant compare to those of natives in the host country? Which segments of the population in the host country benefit or lose from immigration, and how large are these gains and losses?

It is worth noting that our increased understanding of these issues resulted from both theoretical and empirical developments. The *joint* application of economic theory and econometric methods to analyze the many questions raised by immigration has been a

[77] See, for example, Katz and Murphy (1992) and Murphy and Welch (1992).

distinctive feature of recent research in this field, and is mainly responsible for the research advances.

It should not be surprising that in a subject as far-reaching as immigration, there remain many outstanding questions. For example, the economic literature has not devoted sufficient attention to the public finance implications of immigration for the host country. Although many "accounting exercises" in the United States purport to compare the taxes paid by immigrants to the expenditures incurred by governments in the receiving areas, these exercises tend to be purely mechanical and use few insights from the public finance literature. In fact, the link between immigration and the welfare state in many host countries not only raises questions about the tax burden that immigrants might impose on natives, but also about whether the welfare state alters the incentives to migrate and stay in a host country in the first place.

The immigration literature has also downplayed the link between immigration and foreign trade. Economic models suggest that immigration and trade alter national output in the host country by increasing the country's supply of relatively scarce factors of production. As a result, the economic incentives that motivate particular types of workers to migrate to a host country motivate those same workers to produce goods that can be exported to that host country. In the presence of free trade, much of the labor market impact of immigration on the host country would have been observed even in the absence of immigration. A key distinction between immigration and trade, however, is that natives can escape some of the competition from abroad by working in the non-traded sector. Immigrants, however, can move between the traded and non-traded sectors, and natives cannot escape competition from immigrant workers.

The immigration literature has not exploited the fact that different host countries pursue very different immigration policies (and that each country's policy can vary significantly over time). These international differences in immigration policy can be used to evaluate how particular policy parameters influence the labor market impact of immigration on the host country, and may greatly increase our understanding of how immigration alters economic opportunities.

Perhaps the most important topic that has yet to be addressed by the immigration literature concerns the economic impact of immigration on the source country. A relatively large fraction of the population of some source countries has moved elsewhere. Moreover, this emigrant population is not randomly selected, but is composed of workers who have particular sets of skills and attributes. What is the impact of this selective migration on the economic opportunities of those who remain behind? And what is the nature and impact of the economic links that exist between the immigrants in the host country and the remaining population in the source country?

The resurgence of large-scale migration across international boundaries ensures that research in the economics of immigration will continue. The impact of the sizable immigrant flows that have *already* entered many host countries will likely reverberate throughout the host country's economic markets (and social structures) for many decades to come.

As a result, it is unlikely that our interest in the issues raised by the economics of immigration will diminish in the future.

References

Altonji, Joseph G. and David Card (1991), "The effects of immigration on the labor market outcomes of less-skilled natives", in: John M. Abowd and Richard B. Freeman, eds., Immigration, trade and the labor market (University of Chicago Press, Chicago, IL) pp. 201–234.

Baker, Michael and Dwayne Benjamin (1994), "The performance of immigrants in the Canadian labor market", Journal of Labor Economics, 12 (3): 369–405.

Barrett, Alan M. (1993), "Three essays on the labor market characteristics of immigrants", Unpublished PhD dissertation (Michigan State University).

Barro, Robert J. (1991), "Economic growth in a cross-section of countries", Quarterly Journal of Economics 106 (2): 407–433.

Barro, Robert J. and Xavier Sala-i-Martin (1992), "Convergence", Journal of Political Economy 100 (2): 223–251.

Bartel, Ann P. (1989), "Where do the new U.S. immigrants live?" Journal of Labor Economics 7 (4): 371–391.

Beggs, John J. and Bruce J. Chapman (1991), "Male immigrant wage and unemployment experience in Australia", in: John M. Abowd and Richard B. Freeman, eds., Immigration, trade and the labor market (University of Chicago Press, Chicago, IL) pp. 369–384.

Benhabib, Jess (1996), "On the political economy of immigration", European Economic Review 40 (9): 1737–1743.

Ben-Porath, Yoram (1967), "The production of human capital and the life cycle of earnings", Journal of Political Economy 75 (4): 352–365.

Bhagwati, Jagdish N. and T.N. Srinivasan (1983), Lectures on international trade (The MIT Press, Cambridge, MA).

Blanchard, Olivier Jean and Lawrence F. Katz (1992), "Regional evolutions", Brookings Papers on Economic Activity 1: 1–61.

Bloom, David E. and Morley Gunderson (1991), "An analysis of the earnings of Canadian immigrants", in: John M. Abowd and Richard B. Freeman, eds., Immigration, trade and the labor market (University of Chicago Press, Chicago, IL) pp. 321–342.

Bloom, David E., Gilles Grenier and Morley Gunderson (1995), "The changing labour market position of Canadian immigrants", Canadian Journal of Economics 28 (4b): 987–1005.

Borjas, George J. (1983), "The substitutability of black, Hispanic and white labor", Economic Inquiry 21 (1): 93–106.

Borjas, George J. (1985), "Assimilation, changes in cohort quality and the earnings of immigrants", Journal of Labor Economics 3 (4): 463–489.

Borjas, George J. (1987), "Self-selection and the earnings of immigrants", American Economic Review 77 (4): 531–553.

Borjas, George J. (1991), "Immigration and self-selection", in: John M. Abowd and Richard B. Freeman, eds., Immigration, trade and the labor market (University of Chicago Press, Chicago, IL) pp. 29–76.

Borjas, George J. (1992), "Ethnic capital and intergenerational mobility", Quarterly Journal of Economics 107 (1): 123–150.

Borjas, George J. (1994), "The economics of immigration", Journal of Economic Literature 32 (4): 1667–1717.

Borjas, George J. (1995a), "Assimilation and changes in cohort quality revisited: what happened to immigrant earnings in the 1980s?" Journal of Labor Economics 13 (2): 201–245.

Borjas, George J. (1995b), "The economic benefits from immigration", Journal of Economic Perspectives 9 (2): 3–22.

Borjas, George J. (1999), "The economic progress of immigrants", in: George J. Borjas, ed. Issues in the economics of immigration (University of Chicago Press, Chicago, IL) in press.

Borjas, George J. and Bernt Bratsberg (1996), "Who leaves? The outmigration of the foreign-born", Review of Economics and Statistics 78 (1): 165–176.

Borjas, George J. and Stephen G. Bronars (1991), "Immigration and the family", Journal of Labor Economics 9 (2): 123–148.

Borjas, George J., Stephen G. Bronars and Stephen J. Trejo (1992), "Self-selection and internal migration in the United States", Journal of Urban Economics 32 (2): 159–185.

Borjas, George J., Richard B. Freeman and Lawrence F. Katz (1992), "On the labor market impacts of immigration and trade", in: George J. Borjas and Richard B. Freeman, eds., Immigration and the work force: economic consequences for the United States and source areas (University of Chicago Press, Chicago, IL) pp. 213–244.

Borjas, George J., Richard B. Freeman and Lawrence F. Katz (1997), "How much do immigration and trade affect labor market outcomes", Brookings Papers on Economic Activity 1: 1–67.

Bratsberg, Bernt (1995), "The incidence of non-return among foreign students in the United States", Economics of Education Review 14 (4): 373–384.

Butcher, Kristin F. and John DiNardo (1996), "The immigrant and native-born wage distributions: evidence from United States censuses", Unpublished paper (Boston College).

Card, David (1990), "The impact of the Mariel boatlift on the Miami labor market", Industrial and Labor Relations Review 43 (2): 245–257.

Card, David (1997), "Immigrant inflows, native outflows and the local labor market impacts of higher immigration", Working paper no. 5927 (NBER, Cambridge, MA).

Carliner, Geoffrey (1980), "Wages, earnings and hours of first, second and third generation American males", Economic Inquiry 18 (1): 87–102.

Carrington, William J. and Pedro de Lima (1996), "The impact of 1970s repatriates from Africa on the Portuguese labor market", Industrial and Labor Relations Review 49 (2): 330–347.

Chiswick, Barry R. (1978), "The effect of Americanization on the earnings of foreign-born men", Journal of Political Economy 86 (5): 897–921.

Chiswick, Barry R. (1986), "Is the new immigration less skilled than the old?" Journal of Labor Economics 4 (2): 168–192.

Cobb-Clark, Deborah A. (1983), "Immigrant selectivity and wages: the evidence for women", American Economic Review 83 (4): 986–993.

Dahl, Gordon B. (1997), "Mobility and the returns to education: testing a Roy model with multiple markets", Unpublished paper (Princeton University, Princeton, NJ).

De New, John P. and Klaus F. Zimmermann (1994), "Native wage impacts of foreign labor: a random effects panel analysis", Journal of Population Economics 7 (2): 177–192.

DiNardo, John (1997), "Comments and discussion", Brookings Papers on Economic Activity 1: 68–76.

Douglas, Paul (1919), "Is the new immigration more unskilled than the old?" Journal of the American Statistical Association 16: 393–403.

Duleep, Harriet Orcutt and Mark C. Regets (1996), "Earnings convergence: does it matter where immigrants come from or why", Canadian Journal of Economics 29: S130–S134.

Duleep, Harriet Orcutt and Mark C. Regets (1997), "Are lower immigrant earnings at entry associated with faster growth? A review," Unpublished paper (The Urban Institute).

Dustmann, Christian (1993), "Earnings adjustment of temporary immigrants", Journal of Population Economics 6 (2): 153–168.

Feldstein, Martin S. and Charles Horioka (1980), "Domestic savings and international capital flows", Economic Journal 90 (358): 314–329.

Filer, Randall K. (1992), "The impact of immigrant arrivals on migratory patterns of native workers", in: George J. Borjas and Richard B. Freeman, eds., Immigration and the work force: economic consequences for the United States and source areas (University of Chicago Press, Chicago, IL) pp. 245–269.

Freeman, Richard B. (1977), "Manpower requirements and substitution analysis of labor skills: a synthesis", Research in Labor Economics 1: 151–183.

Frey, William (1995), "Immigration and internal migration 'flight' from US metropolitan areas: toward a new demographic balkanization", Urban Studies 32 (4–5): 733–757.

Frey, William and Kao-Lee Liaw (1996), "The impact of immigration on population redistribution within the United States", Unpublished paper (Population Studies Center Research, University of Michigan).

Friedberg, Rachel M. (1992), The labor market assimilation of immigrants in the United States: the role of age at arrival (Brown University).

Friedberg, Rachel and Jennifer Hunt (1995), "The impact of immigration on host county wages, employment and growth", Journal of Economic Perspectives 9 (2): 23–44.

Funkhouser, Edward and Stephen J. Trejo (1995), "The decline in immigrant labor market skills: did it continue in the 1980s?" Industrial and Labor Relations Review 48 (4): 792–811.

Greenwood, Michael J. (1975), "Research on internal migration in the United States: a survey", Journal of Economic Literature 13 (2): 397–433.

Grossman, Jean Baldwin (1982), "The substitutability of natives and immigrants in production", Review of Economics and Statistics 54 (4): 596–603.

Hamermesh, Daniel (1993), Labor demand (Princeton University Press, Princeton, NJ).

Heckman, James J. (1979), "Sample selection bias as a specification error", Econometrica 47 (1): 153–161.

Heckman, James J. and Bo E. Honoré (1990), "The empirical content of the Roy model", Econometrica 58 (5): 1121–1149.

Hicks, John R. (1932), The theory of wages (Macmillan, New York).

Hunt, Jennifer (1992), "The impact of the 1962 repatriates from Algeria on the French labor market", Industrial and Labor Relations Review 45 (3): 556–572.

Jaeger, David A. (1996), "Skill differences and the effect of immigrants on the wages of natives", Unpublished paper (US Bureau of Labor Statistics).

Jasso, Guillermina and Mark R. Rosenzweig (1986), "What's in a name? country-of-origin influences on the earnings of immigrants in the United States", Research in Human Capital and Development 4: 75–106.

Johnson, George E. (1970), "The demand for labor by educational category", Southern Economic Journal 37 (2): 190–204.

Johnson, George E. (1997), "Estimation of the impact of immigration on the distribution of income among minorities and others", Unpublished paper (University of Michigan).

Katz, Lawrence F. and Kevin M. Murphy (1992), "Changes in the wage structure, 1963–87: supply and demand factors", Quarterly Journal of Economics 107 (1): 35–78.

LaLonde, Robert J. and Robert H. Topel (1991), "Labor market adjustments to increased immigration", in: John M. Abowd and Richard B. Freeman, eds., Immigration, trade and the labor market (University of Chicago Press, Chicago, IL) pp. 167–199.

LaLonde, Robert J. and Robert H. Topel (1992), "The assimilation of immigrants in the U.S. labor market", in: George J. Borjas and Richard B. Freeman, eds., Immigration and the work force: economic consequences for the United States and source areas (University of Chicago Press, Chicago, IL) pp. 67–92.

LaLonde, Robert J. and Robert H. Topel (1996), "Economic impact of international migration and the economic performance of immigrants", in: Mark R. Rosenzweig and Oded Stark, eds., Handbook of population and family economics (North-Holland, Amsterdam).

Martin, Phillip (1998), Migration news (University of California, Davis, CA).

Mincer, Jacob (1974), Schooling, experience and earnings (Columbia University Press, New York).

Mincer, Jacob (1978), "Family migration decisions", Journal of Political Economy 86 (5): 749–773.

Murphy, Kevin M. and Finis Welch (1992), "The structure of wages", Quarterly Journal of Economics 107 (1): 215–326.

National Research Council (1997), The new Americans: economic, demographic and fiscal effects of immigration (National Academy Press, Washington, DC).

Pischke, Jörn-Steffen (1993), "Assimilation and the earnings of guestworkers in Germany", Unpublished paper (MIT, Boston, MA).

Pischke, Jörn-Steffen and Johannes Velling (1997), "Employment effects of immigration to Germany: an analysis based on local labor markets", Review of Economics and Statistics 79 (4): 594–604.

Roy, Andrew D. (1951), "Some thoughts on the distribution of earnings", Oxford Economic Papers 3: 135–146.

Schoeni, Robert F. (1997), "The effect of immigrants on the employment and wages of native workers: evidence from the 1970s and 1980s", Unpublished paper (RAND, Santa Monica, CA).

Sjaastad, Larry A. (1962), "The costs and returns of human migration", Journal of Political Economy 70 (Supplement): 80–93.

Taylor, J. Edward (1987), "Undocumented Mexico-U.S. migration and the returns to households in rural Mexico", American Journal of Agricultural Economics 69 (3): 626–638.

Trefler, Daniel (1997), "Immigrants and natives in general equilibrium trade models", Working paper no. 6209 (NBER, Cambridge, MA).

United Nations (1989), World population at the turn of the century (United Nations, New York).

Welch, Finis (1969), "Linear synthesis of skill distribution", Journal of Human Resources 4 (3): 311–327.

Welch, Finis (1979), "Effects of cohort size on earnings: the baby boom babies' financial bust", Journal of Political Economy 87 (5, Part 2): S65–S97.

White, Michael J. and Zai Liang Zai (1993), "The effect of immigration on the internal migration of the native-born population, 1981–90", Unpublished paper (Brown University).

Wright, Richard A., Mark Ellis and Michael Reibel (1997), "The linkage between immigration and internal migration in large metropolitan areas in the United States", Economic Geography 73 (2): 234–254.

Yuengert, Andrew (1994), "Immigrant earnings, relative to what? The importance of earnings function specification and comparison points", Journal of Applied Econometrics 9 (1): 71–90.

Chapter 29

INTERGENERATIONAL MOBILITY IN THE LABOR MARKET

GARY SOLON*

University of Michigan

Contents

Abstract 1762
JEL codes 1762
1 Introduction 1762
2 A simple theoretical model 1763
3 Sibling correlations in earnings 1766
3.1 Statistical model 1767
3.2 Empirical studies 1769
3.3 What we have learned and what we still do not know 1775
4 Intergenerational correlations in earnings 1776
4.1 Statistical model 1776
4.2 Empirical studies 1778
4.3 What we have learned and what we still do not know 1789
5 Neighborhood effects 1790
5.1 Statistical model 1790
5.2 Empirical studies 1791
5.3 What we have learned and what we still do not know 1794
6 Conclusions 1795
References 1796

* Many of the ideas in this chapter originated during collaborations with Laura Nelson Chadwick, Mary Corcoran, Greg Duncan, Roger Gordon, Deborah Laren, and Marianne Page, and I thank them all. I am grateful to Anders Bjorklund, John Bound, Miles Corak, Julie Berry Cullen, Arthur Goldberger, Susan Mayer, Casey Mulligan, Jerry Solon, Laurie Solon, and Robert Willis for commenting on an earlier draft of this chapter.

Handbook of Labor Economics, Volume 3. Edited by O. Ashenfelter and D. Card

Abstract

This chapter summarizes what has been learned from recent research on intergenerational transmission of earnings status. The chapter begins by using a simple theoretical model to highlight several key concepts. Then it reviews (and discusses the connections among) three related empirical literatures: on sibling correlations in earnings, on the intergenerational elasticity of offspring's earnings with respect to parents' earnings or income, and on neighborhood effects.

JEL codes: D1; D3; J3

1. Introduction

Imagine two societies, society A and society B. The distribution of earnings is identical between these two societies, so no matter how one measures inequality – using the variance of log earnings, the Gini coefficient, or whatever – one finds that the degree of cross-sectional inequality is the same in both societies. At first glance, the two societies appear to be equally unequal. But now suppose that, in society A, one's relative position in the earnings distribution is exactly inherited from one's parents. If your parents were in the 90th percentile of earnings in their generation, it is certain that you place in the 90th percentile in your own generation. If your parents were in the 5th percentile in their generation, you also inevitably place in the 5th percentile. As far as earnings are concerned, society A is an extreme caste society. In contrast, in society B, one's relative position in the earnings distribution is completely independent of the position of one's parents. The offspring of parents in the 5th percentile and the offspring of parents in the 90th percentile show the same distribution of earnings. Unlike society A, society B displays complete intergenerational mobility.

Although societies A and B have the same measured inequality within a generation, the two societies are tremendously different in the *character* of their inequality. And, once that is agreed to, we should wonder where our own society's intergenerational mobility lies along the spectrum between societies A and B, and how it compares to the mobility in other societies. Beyond that, we would like to know *why* each society has the degree of mobility it does. Even with all that information, reasonable people still would disagree about the fairness of their society's degree of mobility and about what, if anything, should be done about it. But without such information, it is difficult to have an informed opinion.

The purpose of this chapter is to summarize what has been learned from recent research on intergenerational mobility. Because this is a chapter in a handbook of *labor* economics, I will focus mainly on the intergenerational transmission of labor earnings. I will not discuss the intergenerational transmission of other wealth. I also will mostly overlook some literatures that clearly are related to intergenerational mobility in the labor market. I will not cover the vast sociology literature on intergenerational mobility across occupational categories, which has been reviewed elsewhere by Erikson and Goldthorpe (1992)

and Ganzeboom et al. (1991). And, although I will give some attention to the role of neighborhood background in intergenerational mobility, I will not discuss the closely related literature on school effects, recent reviews of which include Willis (1986), Card (1995), Hanushek (1986), Betts (1996), and Card and Krueger (1996).[1]

That still leaves plenty to talk about. In the next section, I will present a variant of the Becker and Tomes (1979) model of intergenerational mobility and use it to highlight several key concepts. In Section 3, I will review the literature that uses sibling correlations as omnibus measures of the overall influence of family and community background on earnings. In Section 4, I will review the rapidly expanding empirical literature on the association between children's and parents' earnings. In Section 5, I will discuss the recent empirical literature on neighborhood effects. Section 6 will comment on what we have learned so far, what we still do not know, and how we might find out more.

2. A simple theoretical model

The focus of the subsequent sections of this chapter will be primarily empirical. Surveys of the theoretical literature on intergenerational mobility can be found elsewhere in Mulligan (1997) and Behrman (1997). The interpretation of the empirical evidence reviewed in this chapter will be enhanced, however, by considering the following simplified version of the theoretical model in Becker and Tomes (1979).

A family containing one parent and one child must allocate the parent's lifetime earnings y_{t-1} between the parent's own consumption C_{t-1} and investment I_{t-1} in the child's earning capacity. Thus, the budget constraint is

$$y_{t-1} = C_{t-1} + I_{t-1}. \tag{1}$$

The technology translating the investment I_{t-1} into the child's lifetime earnings y_t is

$$y_t = (1 + r)I_{t-1} + E_t, \tag{2}$$

where r is a parametric return to human capital investment and E_t represents the combined effect of all other determinants of the child's lifetime earnings.

The family's decision-maker (the parent?) divides y_{t-1} between C_{t-1} and I_{t-1} to maximize the Cobb–Douglas utility function

$$U = (1 - \alpha)\log C_{t-1} + \alpha \log y_t, \tag{3}$$

where knowledge of E_t is assumed and the parameter α, which lies between 0 and 1, indexes the decision-maker's taste for y_t relative to C_{t-1}. The first-order conditions for this maximization imply that the optimal choice of I_{t-1} is

$$I_{t-1} = \alpha y_{t-1} - (1 - \alpha)E_t/(1 + r). \tag{4}$$

[1] I also will not discuss the literatures on intergenerational associations in welfare program participation (e.g., Solon et al., 1988) and in consumption expenditures on food and housing (Aughinbaugh, 1996; Mulligan, 1997).

Then substituting Eq. (4) in for I_{t-1} in Eq. (2) yields

$$y_t = \beta y_{t-1} + \alpha E_t, \tag{5}$$

where $\beta = \alpha(1 + r)$.

A first glance at Eq. (5) gives the impression that, if the process for y is stationary so that the variance of y is the same in each generation, then β is the correlation between the child's and parent's lifetime earnings. But that requires that E_t be orthogonal to y_{t-1}, which it generally is not. To explain why it is not, I follow Becker and Tomes in decomposing E_t as

$$E_t = e_t + u_t, \tag{6}$$

where e_t is the child's "endowment" of earning capacity (aside from the part resulting from the parent's conscious investment I_{t-1}) and u_t is the child's "market luck," assumed to be independent of y_{t-1} and e_t. The endowment e_t represents the combined effect of many child attributes influenced by nature, nurture, or both. In Becker and Tomes' words, children's endowments "are determined by the reputation and 'connections' of their families, the contribution to the ability, race, and other characteristics of children from the genetic constitutions of their families, and the learning, skills, goals, and other 'family commodities' acquired through belonging to a particular family culture. Obviously, endowments depend on many characteristics of parents, grandparents, and other family members and may also be culturally influenced by other families."

With this characterization of the sources of the endowment, it is natural to assume that the child's endowment e_t is positively correlated with the parent's endowment e_{t-1}. Becker and Tomes assume, in particular, that e_t follows the first-order autoregressive process

$$e_t = \lambda e_{t-1} + v_t, \tag{7}$$

where $0 \leq \lambda < 1$, v_t is serially uncorrelated with variance σ_v^2, and from here on I suppress intercepts by expressing all variables in deviation-from-mean form. Returning to Eq. (5), it now is clear that, as long as λ is positive, E_t is positively correlated with y_{t-1} because both depend on the parent's endowment e_{t-1}, and hence the intergenerational earnings correlation is not simply β.

To ascertain what the intergenerational earnings correlation is, substitute Eq. (6) into Eq. (5) to get

$$y_t = \beta y_{t-1} + \alpha e_t + \alpha u_t, \tag{8}$$

and assure stationarity in the process for y by assuming that $0 < \beta < 1$, the population variance of e_t is $\sigma_e^2 = \sigma_v^2/(1 - \lambda^2)$ for all t, and the population variance of u_t is σ_u^2 for all t. Then the intergenerational correlation between y_t and y_{t-1} becomes immediately apparent for two familiar special cases. First, if $\sigma_e^2 = 0$ or $\lambda = 0$, then Eq. (8) is just a first-order autoregressive process for y with a white-noise error term, and the autoregressive parameter β is the intergenerational earnings correlation after all. Second, if $\sigma_u^2 = 0$, then Eq.

(8) is a first-order autoregressive process for y with a first-order autoregressive error term. This situation frequently appears in econometrics textbooks[2] as an example in which ordinary least squares estimation of an autoregressive coefficient is inconsistent. In this situation, the probability limit of the ordinary least squares estimator, which is also the correlation between y_t and y_{t-1}, is $(\beta + \lambda)/(1 + \beta\lambda)$. This correlation exceeds β if $\lambda > 0$ (and symmetrically exceeds λ if $\beta > 0$).

More generally, the intergenerational earnings correlation generated by the model in Eq. (8) is a weighted average of the intergenerational correlations in these two special cases. In particular,

$$\text{Corr}(y_t, y_{t-1}) = \delta\beta + (1 - \delta)[(\beta + \lambda)/(1 + \beta\lambda)], \tag{9}$$

where

$$\delta = \alpha^2\sigma_u^2/[(1 - \beta^2)\sigma_y^2] \tag{10}$$

is the proportion of the variance in y originating from innovations in the u series rather than in the v series. Thus, in the first special case above, when $\sigma_v^2 = 0$ and hence $\sigma_\varepsilon^2 = 0$, all the weight goes on the first term on the right side of Eq. (9), and the intergenerational earnings correlation is β. In the second special case, when $\sigma_u^2 = 0$, all the weight goes on the second term, and the intergenerational correlation is $(\beta + \lambda)/(1 + \beta\lambda)$. When both sources of variance are present, the intergenerational correlation is a weighted average of β and $(\beta + \lambda)/(1 + \beta\lambda)$.

Before discussing what lessons can be drawn from these results, I should emphasize some obvious limitations of this analysis. It ignores the intergenerational transfer of assets other than human capital.[3] It makes some very arbitrary functional form assumptions, such as the form of the utility function in Eq. (3). By assuming single-parent families, it ignores the role of assortative mating in intergenerational mobility.[4] By assuming single-child families, it ignores the role of the division of family resources among multiple children, as well as the effects of interactions among the children.[5]

Despite its extreme simplicity, however, the model still is rich enough to illustrate several crucial aspects of the intergenerational transmission of earnings status. First, even with all that is left out of the model, the model shows that intergenerational transmission occurs through a multitude of processes.[6] Eq. (2) shows that the child's earnings depend partly on investment in the child's human capital, and Eq. (4) shows that the amount of that investment depends partly on parental earnings. Eqs. (2) and (6) show that the child's earnings also depend partly on the child's endowed capacities, which Eq.

[2] See Greene (1997, pp. 586–587) for example.

[3] The distinction between intergenerational transfers of assets and of human capital is a major focus of the analysis in Becker and Tomes (1986).

[4] See Weiss (1997) for a survey of economic models of the marriage "market."

[5] See Behrman (1997).

[6] The difficulty of empirically disentangling the various processes is a major message of Goldberger (1989).

(7) says are influenced – through some combination of nature and nurture – by the parent's endowment.[7] Eq. (9) says that these processes contribute to an intergenerational earnings correlation that depends on numerous parameters, which in turn depend on still other parameters in earlier equations. In particular, the degree of intergenerational mobility depends on the importance that family decision-makers place on the children's future earnings, the return to human capital investment, the strength of the intergenerational transmission of endowments, and the relative magnitudes of the variances in market luck and endowment luck.

Second, although there are many good reasons to expect a positive intergenerational earnings correlation, the correlation need not be large. As Eq. (9) shows, the intergenerational correlation depends on numerous parameters, and theory does not say a great deal about how large or small we should expect most of those parameters to be. To get a clearer notion of how much intergenerational mobility there really is, we need to look at the empirical evidence, which is what we will do in Section 4 of this chapter.

Third, the intergenerational influences on the child's earnings may depend on other aspects of family background besides parental income. The child's earnings depend on the child's endowed earning capacity as well as the parent's earnings, and the child's endowment is partly inherited from the parent's endowment. While the parent's endowment is correlated with the parent's earnings, the correlation is imperfect. It therefore is possible, for example, for the children of low-earning immigrants to inherit talents or cultural values that enable the children to achieve high earnings. A comparison of the empirical evidence on sibling correlations summarized in Section 3 with the evidence on intergenerational correlations summarized in Section 4 will suggest that a large share of whatever it is about family and community background that affects children's earnings is indeed uncorrelated with parental income. What these mysterious background factors are is one of the major questions to be addressed by future research.

3. Sibling correlations in earnings

Numerous researchers have used sibling correlations in socioeconomic outcomes to measure the proportion of the variation in those outcomes that can be attributed to family and community background variables (including unmeasured ones). The basic idea is that, if family and community origins play a large role in determining socioeconomic status, siblings will show a strong resemblance in their status; if family and community background matters hardly at all, siblings will show little more resemblance than would randomly selected unrelated individuals. The first part of this section will use a simple statistical model to formalize this idea and also to illustrate some problems in the estimation of sibling correlations. The second part will review the empirical evidence on sibling

[7] This avenue for intergenerational transmission has been emphasized recently by Herrnstein and Murray (1994), Mayer (1997), and Shea (1997).

correlations in earnings, and the third part will summarize what has been learned so far and what remains to be studied.

3.1. Statistical model[8]

Let y_{ij} be some measure of the long-run earnings (for example, the permanent component of log annual earnings) of the jth sibling in family i. A simple way of characterizing the role of family and community background is to assume that y_{ij} can be additively decomposed as

$$y_{ij} = a_i + b_{ij}, \tag{11}$$

where the family component a_i represents the combined effect of all factors common to siblings from family i and the orthogonal sibling-specific component b_{ij} denotes the combined effect of all factors purely idiosyncratic to sibling j. Then, letting σ_y^2, σ_a^2, and σ_b^2 denote the respective population variances of y_{ij}, a_i, and b_{ij}, the population variance in long-run earnings is the sum of the two sources of variation:

$$\sigma_y^2 = \sigma_a^2 + \sigma_b^2. \tag{12}$$

The covariance in long-run earnings between siblings j and j' from the same family,

$$\mathrm{Cov}(y_{ij}, y_{ij'}) = \sigma_a^2, \tag{13}$$

identifies the variance component arising from factors shared by siblings.

Then the sibling correlation,

$$\mathrm{Corr}(y_{ij}, y_{ij'}) = \mathrm{Cov}(y_{ij}, y_{ij'})/\sigma_y^2 = \sigma_a^2/(\sigma_a^2 + \sigma_b^2) \tag{14}$$

measures the proportion of the variance in long-run earnings due to whatever factors are shared by siblings. In that sense, the sibling correlation is an index of the extent to which permanent earnings inequality arises from disparities in family and community background. In some respects the sibling correlation is a broad measure of the role of family and community origins, and in other respects it is a narrow one. On one hand, *anything* shared by siblings contributes to the sibling correlation. These shared factors include not only parental socioeconomic status, but also other parental characteristics, the number of children in the family, those interactions among the children that induce sibling resemblance, and shared community factors such as school quality and socioeconomic status of neighbors. On the other hand, some factors sometimes thought of as background factors are left out. For example, those genetic traits *not* shared by siblings are excluded. Also excluded are family or community factors that differ among siblings because the siblings are raised at different times, because the parents treat the siblings differently, because the siblings strive to differentiate themselves from each other, or because of other birth-order

[8] This subsection draws heavily from Solon et al. (1991).

effects. Nonetheless, sibling correlations have considerable appeal as rough omnibus measures of the importance of family and community background, and numerous researchers have used them as such. The next subsection will summarize the results from the many empirical studies that have estimated sibling correlations in earnings.

Before proceeding to that review, however, I will use this subsection's statistical model to highlight some chronic difficulties in the estimation of sibling correlations in earnings. To begin with, although we are mainly interested in $\text{Corr}(y_{ij}, y_{ij'})$, the sibling correlation in *long-run* earnings, most of the empirical literature has estimated brother correlations in *single-year* measures of earnings. The "noisiness" of single-year earnings as an indicator of long-run earnings causes an attenuation inconsistency in the estimation of $\text{Corr}(y_{ij}, y_{ij'})$ similar to the textbook errors-in-variables inconsistency in least squares estimation of the slope coefficient in a simple regression.

To formalize this point, suppose that the available earnings variable for sibling j in family i is y_{ijt}, his log earnings in year t, and suppose that the relationship between this available measure and the desired variable y_{ij} is

$$y_{ijt} = y_{ij} + w_{ijt}, \tag{15}$$

where $\text{Var}(w_{ijt}) = \sigma_t^2$ and $\text{Cov}(y_{ij}, w_{ijt}) = 0$. The measurement error w_{ijt} in current log earnings as an indicator of permanent status arises both from response error in reporting of current earnings and from true transitory fluctuations in current earnings around their longer-run tendency. (For simplicity, I presently am ignoring the additional discrepancy between y_{ijt} and y_{ij} due to the tendency for annual earnings to grow with work experience. As shown in the next subsection, that tendency is readily accounted for by regression adjustments for experience or age.)

What then is the connection between the commonly estimated sibling correlation in the log of single-year earnings and the sibling correlation in the permanent component of log earnings? Suppose that the sibling correlation in the measurement error w_{ijt} is approximately zero, as indicated by some evidence discussed in Solon et al. (1991, footnote 2). Then, since $\text{Var}(y_{ijt}) = \sigma_a^2 + \sigma_b^2 + \sigma_t^2$ and $\text{Cov}(y_{ijt}, y_{ij't}) \cong \sigma_a^2$, the sibling correlation in the log of single-year earnings is

$$\begin{aligned}\text{Corr}(y_{ijt}, y_{ij't}) &\cong \sigma_a^2/(\sigma_a^2 + \sigma_b^2 + \sigma_t^2) = [\sigma_a^2/(\sigma_a^2 + \sigma_b^2)][(\sigma_a^2 + \sigma_b^2)/(\sigma_a^2 + \sigma_b^2 + \sigma_t^2)] \\ &= \text{Corr}(y_{ij}, y_{ij'})[\text{Var}(y_{ij})/\text{Var}(y_{ijt})] \end{aligned} \tag{16}$$

That is, the sibling correlation in the log of single-year earnings approximately equals the sibling correlation in the permanent component attenuated by a factor equal to the proportion of the cross-sectional variance in log annual earnings that is due to variation in the permanent component.

Fortunately, we know something about the magnitude of that attenuation factor. A long history of earnings dynamics studies based on longitudinal data[9] suggests that the permanent component's share of the cross-sectional population variance in log annual earnings is somewhere between about 0.5 and 0.7. Thus, as a measure of the sibling correlation in

the permanent component, the sibling correlation in the log of single-year earnings is biased downward by about 30–50%. To put it another way, if we knew the sibling correlation in the log of single-year earnings, we should multiply it by a factor between 1.4 and 2.0 to infer the sibling correlation in the permanent component of log earnings.

In fact, however, we do not know the population value of the sibling correlation in log single-year earnings. Instead, we have estimates based on various samples, and some of these samples are peculiarly homogeneous. Kearl and Pope's (1986) study, for example, uses data on Mormon brothers in nineteenth-century Utah. The family and community background in such a subpopulation is presumably more homogeneous than in the overall US population (or any other general population of interest), so that the subpopulation's variance in the a_i component of Eq. (11) is less than the population variance σ_a^2. As is clear from Eq. (14), unless for some reason the subpopulation's within-family variance also is less than the population-wide σ_b^2, the sibling correlation in the subpopulation will tend to be less than that in the broader population. In effect, in a homogeneous sample in which even unrelated individuals resemble each other, the resemblance among siblings will seem less striking.

In some other sibling studies – such as the Behrman et al. (1977) study of white twin pairs in which both twins served in the armed forces and then survived until, and cooperated with, a succession of surveys – the samples appear to be relatively homogenous with respect to permanent earnings y_{ij}, but it is not clear which variance component, σ_a^2 or σ_b^2, is more severely understated. Even if they are understated in the same proportion, unless there is somehow a corresponding understatement of the measurement error variance σ_t^2, the homogeneity of the sample aggravates the errors-in-variables inconsistency by reducing the attenuation factor in Eq. (16). The problem is that the sample's homogeneity decreases the "signal" without a commensurate decrease in the "noise."

The upshot is that, when the sibling correlation in the log of single-year earnings is estimated with an unrepresentatively homogeneous sample, multiplying the estimate by a correction factor of 1.4–2.0 may be too small of a correction for estimating the sibling correlation in the permanent component of log earnings. With that in mind, let's proceed to a review of the empirical studies in this literature.

3.2. Empirical studies

Most of the many empirical studies of sibling correlations in earnings estimate correlations among American brothers in single-year measures of their earnings. The results of such studies are summarized in Table 1. Perhaps the first thing to notice is that the estimates are quite dispersed, ranging from 0.11 to 0.44. This variability of the estimates should not be

[9] See, for example, Bowles' (1972) discussion based on evidence in Friedman (1957) as well as more recent studies by Lillard and Willis (1978), Gordon (1984), Solon et al. (1991), Bjorklund (1993), Baker (1997), Haider (1997), and Baker and Solon (1997). To my knowledge, Bowles' study was the first to stress the empirical importance of the errors-in-variables problem in measuring intergenerational mobility.

Table 1
Estimates of US brother correlations in single-year earnings

Study	Sample source	No. of families in sample[a]	Age range	Variable	Estimated brother correlation
Bound et al. (1986)	NLS	>276[b]	27[c]	Residual from regression of log hourly wage at about age 27 on age, race, and location variables	0.11
Brittain (1977)	Sons of 1964–1965 decedents in Cleveland	66	42[d]	Log annual family income in 1965–66	0.44
Chamberlain and Griliches (1975)	Gorseline sample of Indiana brothers with differing educational attainments	156	18–72	Log annual income in 1927	0.37
Corcoran and Datcher (1981)	PSID	206	22–32	Log annual earnings in 1978	0.26
Corcoran and Jencks (1979)	NORC	150	24–63	Log annual earnings in 1972–1973	0.13
	Project Talent	99	28–29	Log hourly wage in 1971–1972	0.21
Griliches (1979)	NLS	247	21–31	Log hourly wage in 1973	0.31
Hauser and Sewell (1986)	Wisconsin high school seniors in 1957 and their brothers	532	20–56	Annual earnings in 1974 or 1976	0.30
Kearl and Pope (1986)	Mormons in 19th century Utah	>410[e]	Not reported	Residual from regression of log annual family income in various years on year and age variables	0.20
Olneck (1977)	Kalamazoo sixth-graders from 1928 to 1950	346	35–59	Log annual earnings in 1973	0.21
Solon et al. (1991)	PSID	135	24–31	Residual from regression of log annual earnings in 1982 on age dummies	0.34

[a] Some studies report number of brothers pairs instead of number of families. This may overstate the number of families in samples containing more than two brothers in some families.

[b] Bound et al. (1986) report sample sizes ranging from 276 to 611 for different variables in their sample correlation matrix. Other information in their paper suggests that the sample size for their wage variable is toward the lower end of that range.

[c] Bound et al. (1986) report a sample mean age of 27.1 with standard deviation 1.5.

[d] Brittain (1977) reports that the median age in his sample is 42, the standard deviation is less than 10, and only a few individuals are under 25 or over 65.

[e] Kearl and Pope (1986) report sample sizes by year that range from 59 in 1855 to 410 in 1900. Because they do not report the extent to which sample families reappear across years, it is unclear how many families are contained in the pooled sample.

too surprising given the small sample sizes on which many of the estimates are based.

The central tendency of the estimates seems to be about 0.25 or a little higher. Somewhat surprisingly, this remains true even if we restrict our attention to estimates based on nationally representative samples. If we multiply this central tendency by a correction factor between 1.4 and 2.0, we conclude that the correlation among American brothers in the permanent component of their log earnings may be about 0.4 or somewhat higher.

This back-of-the-envelope calculation can be compared to the results of a few recent studies that have used longitudinal data on brothers' earnings in attempts to estimate the brother correlation in longer-run earnings measures, instead of single-year measures. Using longitudinal brothers data from the Panel Study of Income Dynamics (PSID), Solon et al. (1991) first estimate the regression of log annual earnings on dummy variables for year and age to adjust for time and lifecycle effects. Then they apply analysis-of-variance procedures to the "residualized" log earnings data to decompose the variance into the components associated with the permanent family and community background factor a_i, the permanent brother-specific factor b_{ij}, and the transitory factor w_{ijt}, which is assumed to follow a first-order autoregressive process. Solon et al. estimate brother correlations of 0.45 in the permanent component of log annual earnings, 0.53 in the permanent component of the log of average hourly earnings (i.e., annual earnings divided by annual hours), 0.34 in the permanent component of log family income, and 0.48 in the permanent component of the log of the ratio of the family's income to the poverty line for families of that size and composition. These estimates are imprecise because, with only a few years of longitudinal data, it is difficult to distinguish what is permanent from what is transitory but serially correlated. Nevertheless, the estimates do seem broadly consistent with the back-of-the-envelope calculations above.[10]

Also using longitudinal PSID data, Altonji (1988) estimates brother correlations in multi-year averages of several earnings-related variables. He estimates brother correlations of 0.37 for the log of average hourly earnings, 0.44 for the log of a directly reported hourly wage rate, and 0.37 for the level of family income. These estimates may tend to underestimate the brother correlations in the permanent versions of these variables because, with a young brothers sample, the averages usually are not over very many years, and the transitory component of log earnings is especially volatile at younger ages.[11] As a result, even the multi-year averages may be fairly "noisy" indicators of permanent status.

In another study based instead on the National Longitudinal Surveys (NLS) of labor market experience, Altonji and Dunn (1991) again estimate brother correlations in multi-year averages. They first estimate regressions of the selected annual earnings measures on year dummies and a cubic in age, and then they average the "residualized" earnings

[10] Jantti and Osterbacka (1996) replicate the study by Solon et al. (1991) with longitudinal data on brothers in Finland. They estimate a 0.27 brother correlation in the permanent component of log earnings and conclude that the share of earnings inequality attributable to family and community origins is smaller in Finland than in the United States.

[11] See Gordon (1984), Bjorklund (1993), and Baker and Solon (1997).

observations over all the years available for the individual. Their estimates of the brother correlations for the averaged versions of these variables are 0.32 for log annual earnings, 0.33 for log hourly wage, and 0.30 for log family income. Altonji and Dunn also estimate the brother correlations in the permanent components of these variables by using a method-of-moments estimation procedure built on the strong assumption that the transitory component w_{ijt} is serially uncorrelated at lags of greater than 2 years. They obtain estimates of 0.37 for log annual earnings, 0.42 for log hourly wage, and 0.38 for log family income. Also using the NLS, Ashenfelter and Zimmerman (1997) estimate a 0.31 brother correlation in the average of the 1978 and 1981 log hourly wages.

Taken together, these estimates based on longitudinal data seem consistent with the conjecture, based on single-year data, that the correlation among American brothers in the permanent component of their log earnings is somewhere around 0.4. If that is right, about 40% of permanent earnings inequality so measured is attributable to variation in family and community origins, and 60% is due to factors not shared by brothers. It is not clear, of course, whether to emphasize that the cup is 40% full or 60% empty (or is it the other way around?). On one hand, a great deal, perhaps the majority, of inequality in permanent earnings is due to factors other than family and community origins. On the other hand, the finding that whatever American brothers share explains 40% of permanent earnings variation may be quite impressive to empirical labor economists accustomed to small R^2's in their log earnings regressions. Indeed, a major point in Corcoran et al. (1976) is that their estimates of the proportion of earnings variation shared by brothers far exceeds what the authors are able to explain in regressions of log earnings on particular observed family background characteristics. In earlier work, the very limited explanatory power of observable family background characteristics had led Jencks et al. (1972, pp. 7–8) to conclude, "Poverty is not primarily hereditary.... Indeed, there is nearly as much economic inequality among brothers raised in the same homes as in the general population."

Thanks to the many empirical studies of brothers data, we now know that conclusion was too strong. In fact, something about the family and community origins shared by brothers accounts for a substantial share of earnings inequality, but we do not know very much about what that something is. The mystery of what underlies the considerable resemblance between brothers in their long-run earnings remains a fascinating puzzle and should be a priority for continuing research.

Only a few studies have extended the analysis of sibling correlations in earnings to sisters. Using single-year wage data from the NLS, Bound et al. (1986) estimate the regression of log hourly wage on age, race, and location variables and then estimate sibling correlations in the "residualized" wage measure. They estimate the sister-sister correlation at 0.34 and the sister-brother correlation at 0.07. The Solon et al. (1991) study of longitudinal PSID data estimates sister-sister correlations of 0.28 for the permanent component of log family income and 0.51 for the permanent component of the log of the ratio of family income to the poverty line. Altonji and Dunn's (1991) study of longitudinal NLS data estimates sister-sister and sister-brother correlations in multi-year averages of several earnings-related variables. For sister-sister pairs, they estimate corre-

lations of 0.26 for log annual earnings, 0.38 for log hourly wage, and 0.45 for log family income. For sister-brother pairs, they estimate correlations of 0.14 for log annual earnings, 0.27 for log hourly wage, and 0.28 for log family income. They also use their method-of-moments procedure to estimate sister-sister and sister-brother correlations in the permanent components of these variables. They estimate sister-sister correlations of 0.26 for log annual earnings, 0.42 for log hourly wage, and 0.73 (!) for log family income, and they estimate sister-brother correlations of 0.32 for log annual earnings, 0.41 for log hourly wage, and 0.56 for log family income. By and large, the sister correlations seem to be roughly as large as the brother correlations. The most regular exception is that the sister-brother correlation in annual earnings appears to be smaller, presumably because of differences between the genders in labor supply behavior.

All of the studies discussed so far pertain to general siblings samples comprised mostly of non-twins. An important branch of the literature, however, focuses on samples of twins, often identical (monozygotic) twins. The studies that have estimated twin correlations in earnings are summarized in Table 2. One would expect the resemblance between identical twins to exceed the resemblance between other siblings because identical twins share the exact same genetic endowment and probably are treated more alike than other siblings are. That is just what the table shows. Despite many differences across studies, including that the samples are drawn from Australia and Sweden as well as the United States, all the estimated earnings correlations between identical twins lie in a surprisingly narrow range from 0.54 to 0.68. These estimates are much higher than those for general sibling pairs as well as those for fraternal (dizygotic) twins. The latter, which range from 0.30 to 0.46, run higher than the central tendency of Table 1's estimated correlations among non-twin siblings, but are dramatically smaller than the corresponding estimates for identical twins.

Under very strong assumptions, the contrast between the correlations for identical and fraternal twins can be used to infer the relative contributions of nature and nurture to earnings variation. Taubman (1976), for example, assumes that the similarity of genetic endowment between fraternal twins contributes half as much to their earnings correlation as the identical genetic endowment of identical twins contributes to theirs. He also assumes that identical twins experience no more (or less) similarity in environment than fraternal twins do. Under these assumptions, any observed contrast between the identical-twins correlation and the fraternal-twins correlation can be generated only by the greater genetic similarity between identical twins. Therefore, the much greater earnings correlation actually observed for identical twins leads to the conclusion that variation in genetic endowments must be the main source of σ_a^2, the component of earnings variation associated with family and community origins. Goldberger (1979) and others, however, have questioned the assumptions leading to this conclusion. If, for example, identical twins experience more similarity in environment than fraternal twins do, it becomes unclear to what extent their greater earnings correlation arises from their more similar nature or their more similar nurture.

Psychologists and geneticists have gone a few steps further with this approach of contrasting sibling correlations among various sibling types. Looking not at earnings,

Table 2
Estimates of twin correlations in earnings

Study	Sample source	No. of families in sample	Age range	Variable	Estimated twin correlation
Ashenfelter and Krueger (1994)	Twinsburg identical twins	149	37[a]	Log hourly wage in 1991	0.56
	Twinsburg fraternal twins	46	36[b]		0.36
Behrman et al. (1977)	NRC identical male twins	1019	46–56	Log annual earnings in 1973	0.54
	NRC fraternal male twins	917			0.30
Isacsson (1997)	Swedish Twin Registry identical twins	2492	29–67	Three-year average of log annual earnings in 1987, 1990, and 1993	0.68
	STR fraternal twins of same gender	3368			0.46
Miller et al. (1995)	Australian Twin Register identical twins	602	18–64	Log of mean 1985 annual earnings in individual's usual occupation	0.68
	ATR fraternal twins	568			0.32
Rouse (1997)	Twinsburg identical twins	453	38[c]	Log hourly wage in 1991, 1992, 1993, or 1995	0.64

[a] Ashenfelter and Krueger (1994) report a sample mean age of 36.6 with standard deviation 10.4.
[b] Ashenfelter and Krueger (1994) report a sample mean age of 35.6 with standard deviation 8.3.
[c] Rouse (1997) reports a sample mean age of 37.9 with standard deviation 11.5.

but at other outcomes such as IQ and measures of personality traits, they have contrasted the sibling correlations for identical twins reared together, identical twins reared apart, fraternal twins reared together and apart, non-twin siblings reared together and apart, adoptive (biologically unrelated) siblings, and so forth. The findings – summarized in Bouchard et al. (1990), Feldman et al. (1996), Plomin and Daniels (1987), and Plomin and Petrill (1997) – do not yield precise conclusions without strong assumptions, but are striking nonetheless. For example, most estimates of the IQ correlation between identical twins reared apart are about 0.70 or a bit higher. The correlation between identical twins reared together is even higher at about 0.85, suggesting a substantial role for common environment. But the correlation between identical twins reared apart exceeds the correlation between fraternal twins reared together and far exceeds the correlation between non-twin siblings reared together, which in turn exceeds the correlation between adoptive siblings. It is difficult to view these results without concluding that genetic endowment also plays quite an important role in IQ variation.

Of course, IQ and personality traits are only a few of the many characteristics that influence earnings, so it is not obvious what these results imply about the relative roles of nature and nurture in determining earnings variation. But we economists might learn something by following the example of other disciplines and studying correlations among various sibling types in the outcome variables, such as earnings, that are in our own scholarly domain. While the natural experiments afforded by variation in sibling configurations are hardly laboratory experiments, the results from the psychologists' and geneticists' studies are thought-provoking, to say the least. Similar analyses of earnings correlations might generate valuable clues about the factors contributing to earnings inequality.

3.3. What we have learned and what we still do not know

The empirical literature on sibling correlations in earnings, mostly focused on brothers in the United States, suggests that somewhere around 40% of the variance in the permanent component of log earnings is generated by variation in the family and community background factors shared by siblings. This finding indicates that the role of family and community origins in accounting for earnings inequality is quite important and is larger than had been apparent from a superficial look at single-year earnings data on siblings or from earlier research that estimated regressions of earnings on observable family characteristics. Nevertheless, just as large a share of long-run earnings variation is due to factors *not* common to siblings, and the question of what causes so much earnings inequality even within families is an important challenge for further research.

Even the portion that *is* common to siblings is not very well understood. One way to investigate the sources of the sibling resemblance in earnings is to examine the relationships between earnings and particular background characteristics. Along those lines, Section 4 of this chapter will discuss the evidence on the connection between parental income and offspring's earnings, and Section 5 will consider the effects of neighborhood

background. Another possible approach, not yet fully exploited by economists, is to study how the sibling resemblance in earnings varies across different sibling types known to vary in the extent to which they share genes and environments. Although economists have contrasted the earnings correlations of identical and fraternal twins, psychologists and geneticists interested in outcomes other than earnings have demonstrated that studying a richer variety of sibling types, such as identical twins reared apart and adoptive siblings, may generate additional clues about why sibling correlations are as large (and small) as they are.

4. Intergenerational correlations in earnings

According to the evidence on sibling correlations reviewed in Section 3, something about the family one comes from matters a lot for one's position in the earnings distribution. This section will examine the extent to which that something is related to parental income or earnings. In the process, it will address the question raised in the introductory section: Where do the societies we inhabit lie between society A, where one inherits one's exact position in the earnings distribution from one's parents, and society B, where one's position is entirely independent of one's parents' position?

The first part of this section will extend the statistical model introduced in Section 3. The extended model will clarify the connection between sibling correlations and intergenerational correlations, and it also will help illustrate some problems in the estimation of intergenerational correlations. The second part will review the rapidly growing empirical literature on intergenerational earnings mobility, and the third part will summarize what has been learned so far and what remains to be studied.

4.1. Statistical model[12]

Eq. (11) in Section 3 modeled y_{ij}, the long-run earnings of sibling j in family i, as the sum of a family component a_i and an orthogonal sibling-specific component b_{ij}. Now suppose that the family component a_i can be further decomposed as

$$a_i = \rho X_i + z_i, \tag{17}$$

where X_i is some measure of the long-run income of the parents in family i, z_i denotes the combined effect of family background characteristics uncorrelated with parental income, and the intercept of this regression equation is suppressed by expressing all variables as deviations from their population means. Then substituting Eq. (17) in for a_i in Eq. (11) yields

$$y_{ij} = \rho X_i + \varepsilon_{ij}, \tag{18}$$

where $\varepsilon_{ij} = z_i + b_{ij}$ and $\mathrm{Cov}(X_i, \varepsilon_{ij}) = 0$.

[12] This subsection draws heavily from Solon (1992) and Corcoran et al. (1990).

Eq. (18) describes the intergenerational association between child's earnings y_{ij} and parental income or earnings X_i. It is similar to the intergenerational relationships discussed in the theoretical model of Section 2, but now we are denoting parental income by X_i instead of y_{t-1}, with the subscript used to index families instead of generations. Also, whereas y_t and y_{t-1} denoted earnings *levels* in Section 3, in this section I generally will follow the empirical literature in using *logarithmic* earnings measures for y_{ij} and X_i. The regression coefficient ρ in Eq. (18) therefore will represent the *elasticity* of child's long-run earnings with respect to parents' long-run earnings or income. It will provide a parametric answer to questions like, if the parents' long-run earnings are 50% above the average in their generation, what percent above the average should we predict the child's long-run earnings to be in her or his generation? In addition, if the variances in the logarithmic earnings variables are about the same in the child's and parents' generations, then ρ also will approximately equal the intergenerational *correlation* between y_{ij} and X_i.

With this setup, we can derive the connection between the intergenerational association analyzed in this section and the sibling correlation discussed in Section 3. Taking variances of both sides of Eq. (17) yields

$$\sigma_a^2 = \rho^2 \sigma_X^2 + \sigma_z^2, \tag{19}$$

and dividing through by σ_y^2 gives

$$\sigma_a^2/\sigma_y^2 = \mathrm{Corr}(y_{ij}, y_{ij'}) = (\rho^2 \sigma_X^2/\sigma_y^2) + (\sigma_z^2/\sigma_y^2). \tag{20}$$

If inequality in logarithmic earnings is about the same in both generations, so that $\sigma_y^2 \cong \sigma_X^2$, then Eq. (20) simplifies to

$$\mathrm{Corr}(y_{ij}, y_{ij'}) \cong \rho^2 + (\sigma_z^2/\sigma_y^2). \tag{21}$$

This expression decomposes the sibling correlation in long-run earnings into two components – the square of the intergenerational elasticity ρ and a second component due to factors uncorrelated with parental income. Once we review the empirical evidence on the magnitude of ρ, we will use Eq. (21) to consider how much of the sibling correlation in long-run earnings is related to parental income and how much is related to factors uncorrelated with parental income.

Before proceeding to that review, we can use our statistical model again to highlight some estimation problems. To begin with, like most of the empirical literature on sibling correlations in earnings, most of the early studies of intergenerational mobility used single-year measures of earnings. Much as the "noisiness" of single-year earnings measures as proxies for long-run earnings causes an attenuation inconsistency in the estimation of the sibling correlation $\mathrm{Corr}(y_{ij}, y_{ij'})$, it also tends to cause underestimation of the intergenerational association in long-run earnings.

Specifically, suppose that the available earnings variable for a child from family i is y_{it}, his log earnings in year t. Suppose that this is related to y_i, the permanent component of his log earnings, by

$$y_{it} = y_i + w_{it}, \tag{22}$$

where $\text{Var}(w_{it}) = \sigma_t^2$ and $\text{Cov}(y_i, w_{it}) = \text{Cov}(X_i, w_{it}) = 0$. Eq. (22) is the same as Eq. (15) in Section 3 except that now, to reduce notational clutter, we are dropping the sibling-specific subscript j. Similarly, suppose that the single-year measure of parental earnings or income in year s follows

$$X_{is} = X_i + w_{is}, \tag{23}$$

where $\text{Var}(w_{is}) = \sigma_s^2$ and $\text{Cov}(X_i, w_{is}) = \text{Cov}(y_i, w_{is}) = \text{Cov}(w_{it}, w_{is}) = 0$. (Again, for simplicity, I am ignoring for now the gap between current and permanent earnings that arises from earnings growth over the lifecycle.)

Under these assumptions, the intergenerational correlation between the single-year measures y_{it} and X_{is} understates the correlation between the long-run earnings variables y_i and X_i:

$$\text{Corr}(y_{it}, X_{is}) = \text{Corr}(y_i, X_i)[\sigma_y \sigma_X / \sqrt{(\sigma_y^2 + \sigma_t^2)(\sigma_X^2 + \sigma_s^2)}] < \text{Corr}(y_i, X_i). \tag{24}$$

In the special case in which $\sigma_y^2 = \sigma_X^2$ and $\sigma_t^2 = \sigma_s^2$, this is the same result as Eq. (16) for sibling correlations. The result is modified only slightly if instead the intergenerational elasticity ρ in Eq. (18) is estimated by applying least squares to the regression of the "noisy" dependent variable y_{it} on the "noisy" explanatory variable X_{is} in a representative sample. In that case, the resulting estimator $\hat{\rho}$ is subject to the textbook errors-in-variables inconsistency:

$$\text{plim}\, \hat{\rho} = \rho \sigma_X^2 / (\sigma_X^2 + \sigma_s^2) < \rho.$$

As discussed in Section 3, the attenuation factor – the share of the permanent component in the cross-sectional population variance of log annual earnings in the parents' generation – is probably between about 0.5 and 0.7. Since ρ would tend to be underestimated by about 30–50% in a representative sample, consistent estimation would require multiplying the initial estimate by a correction factor between about 1.4 and 2.0.

Matters are even worse if the estimates are based on an unrepresentatively homogeneous sample. The parents in Behrman and Taubman's (1985) study of intergenerational mobility, for example, were drawn from Behrman et al.'s (1977) homogeneous sample of white male twins, in which both members of each twin pair had served in the armed forces and then cooperated with a succession of surveys. Inspection of Eqs. (23) and (24) reveals that the small "signal" variance associated with such a sample worsens the errors-in-variables inconsistency unless somehow the "noise" variance also is commensurately reduced. In such cases, estimates based on single-year earnings data may underestimate the intergenerational association in long-run earnings by *more* than 30–50%.

4.2. Empirical studies

Most of the early studies of intergenerational earnings mobility are surveyed in Section V

of Becker and Tomes (1986). As Becker and Tomes say, "The point estimates for most of the studies indicate that a 10% increase in father's earnings (or income) raises son's earnings by less than 2%." Based on one such study, Behrman and Taubman (1985) had concluded, "The members of this sample come from a highly mobile society," and Becker and Tomes' summary of the evidence is much the same: "Regression to the mean in earnings in rich countries appears to be rapid." Later, in his presidential address to the American Economic Association, Becker (1988) similarly concluded, "In all these countries, low earnings as well as high earnings are not strongly transmitted from fathers to sons...." In other words, these countries appeared more like society B than society A.

Most of the studies leading to these conclusions, however, used single-year earnings or income measures, and, in many, the resulting attenuation inconsistency was aggravated by reliance on peculiarly homogeneous samples. Consequently, during the 1990s, a new wave of intergenerational mobility studies has attempted to reestimate the intergenerational elasticity ρ with new data and methods less susceptible to the estimation problems highlighted in the previous subsection.

Most of the new studies have been based on two US longitudinal surveys, the PSID and the NLS. Both of these surveys began with national probability samples in the late 1960s and then followed the children in the sampled families as they matured into adulthood and formed their own households. In recent years, therefore, it has become possible for researchers to relate the children's earnings status as adults to their parents' status, with both generations' income variables contemporaneously self-reported. The newly available intergenerational data from these two surveys have offered two major analytical advantages over the datasets previously available. First, because the data come from national probability samples, they avoid the homogeneity of the samples used in many early studies. Second, the longitudinal nature of the data enables exploration of the empirical importance of using long-run instead of short-run income measures.

Even if only single-year earnings measures are used, the larger "signal" afforded by the heterogeneity of these samples should reduce the errors-in-variables inconsistency and produce larger estimates of the intergenerational elasticity. As one example, the first columns of Tables 2 and 3 in Solon (1992) show the results from least squares estimation of the regression of the PSID sons' log earnings in 1984 on single-year measures (from each year between 1967 and 1971) of their fathers' log earnings, with age controls to account for both generations' lifecycle variation in annual earnings. Despite the use of father's single-year log earnings as the key regressor, the estimates of the intergenerational earnings elasticity are around 0.3, considerably higher than most of the earlier estimates surveyed by Becker and Tomes.

Even these estimates, though, are presumably subject to an attenuation factor of something like 0.5–0.7 because father's single-year log earnings are a "noisy" indicator of his permanent earnings. Indeed, the remaining columns in Tables 2 and 3 of my 1992 paper show that the estimated intergenerational elasticity does increase as the "noise" in father's log earnings is reduced by averaging his log earnings over progressively more years. The estimate rises to 0.41 once father's log earnings are averaged over all five years between

1967 and 1971. Although the "noise" remaining in even a 5-year average presumably induces at least a minor errors-in-variables inconsistency, this estimate suggests a much less mobile society than depicted by the earlier studies.

Numerous other studies have similarly used the PSID or NLS to estimate regressions of son's log earnings on various parental log income measures along with age controls for both generations. Table 3 summarizes the least squares estimates of the intergenerational elasticity based on the PSID, and Table 4 summarizes those based on the NLS.[13] As usual when comparing empirical studies that differ in many dimensions, the estimates vary considerably,[14] and the variation defies simple explanations. Nevertheless, most of the estimates of the intergenerational elasticity ρ fall in a range between about 0.3 and 0.5.

One systematic pattern that arguably is present in the tables is that the lowest estimates tend to be from samples that include sons observed particularly early in their careers. Such a pattern is a major focus of the study by Reville (1995), which estimates regressions of 5-year averages of son's log earnings on 5-year averages of father's log earnings. When the sons' averages are taken over years when the sons were still in their twenties, the estimates of the intergenerational elasticity are around 0.25. When instead the sons' earnings are averaged over years when the sons were well into their thirties, the elasticity estimates start approaching 0.5. These results suggest that, in the early career years, the measurement error in son's current log earnings as a proxy for longer-run status may not be of the classical variety assumed in the previous subsection. If, among sons in their twenties, the ones destined for higher long-run earnings are about to experience more rapid earnings growth than the ones destined for lower long-run earnings, the measurement error in the earlier years is "mean-reverting" and causes a downward inconsistency in the estimation of the intergenerational elasticity.

Several of the studies listed in Tables 3 and 4 supplement their least squares estimates of the intergenerational elasticity with alternative estimates based on other procedures. For example, as discussed in Section 3, Altonji and Dunn (1991) estimate sibling correlations in the permanent component of various income variables with a method-of-moments procedure built on the strong assumption that the transitory component is serially uncorrelated at lags of greater than 2 years. Applying the same procedure to intergenerational correlations, Altonji and Dunn estimate father-son correlations of 0.39 in the permanent

[13] Several studies are omitted from the tables because they do not fit neatly into the tables' format. Altonji (1988) reports sample correlations between son's and father's log earnings without any age adjustments, and his estimated regressions that do control for age also control for both generations' education. Behrman and Taubman (1990) pool daughters and sons together in their intergenerational regressions. Lillard and Kilburn's (1996) estimates pertain to complex and highly restrictive models of the earnings covariances among various family members. In any case, all these studies, like most of those in Tables 3 and 4, estimate greater intergenerational earnings associations than were typically found in the earlier studies.

[14] As discussed in Buron (1994), Couch and Lillard (1994), and Solon (1994), the estimates vary even more widely if observations of zero earnings are included in the analysis samples. Couch and Lillard estimate negative intergenerational associations in some of their analyses, and the estimates in my comment on their work range from 0.05 to 0.53. The instability of the estimates seems to arise mainly because observations of zero for father's earnings are an example of the "leverage points" discussed in Krasker et al. (1983).

Table 3
Least squares estimates of the elasticity of son's earnings with respect to parental income: PSID

Study	Son's outcome variable	Sons' age range	Measure of parental income	$\hat{\rho}$
Bjorklund and Jantti (1997)	Log annual earnings in 1987	28–36	5-year average of father's log annual earnings	0.39
Buron (1994)	Log of 5-year average of annual earnings	25–37	Log of 5-year average of father's annual earnings	0.39
Couch and Dunn (1997)	Log of multi-year (up to 6-year) average of annual earnings	Not reported[a]	Log of multi-year (up to 6-year) average of father's annual earnings	0.13
Couch and Lillard (1994)	Log annual earnings in 1984	25–33	Log of 5-year average of father's annual earnings	0.53
Eide and Showalter (1997)	Log annual earnings in 1991	24–40	Log of 3-year average of father's annual earnings	0.34
Lillard and Reville (1996)	3-year average of log annual earnings	25–40	3-year average of father's log annual earnings	0.28
Minicozzi (1997)	Log of 2-year average of annual earnings	28–29	Log of estimate of present discounted value of father's lifetime earnings	0.42
			Log of estimate of present discounted value of parents' lifetime earnings	0.53
Mulligan (1997)	Log of multi-year (up to 5-year) average of annual earnings	23–37	Log of multi-year (up to 5-year) average of father's annual earnings	0.32
	Log of multi-year (up to 5-year) average of hourly earnings		Log of multi-year (up to 5-year) average of father's hourly earnings	0.33
	Log of multi-year (up to 5-year) average of family income		Log of multi-year (up to 5-year) average of parents' family income	0.48

Table 3 (*continued*)

Study	Son's outcome variable	Sons' age range	Measure of parental income	$\hat{\rho}$
Reville (1995)	5-year average of son's log annual earnings	32–40	5-year average of father's log annual earnings	0.48
Shea (1997)	Log of multi-year (on average 8-year) average of annual earnings	25–40	Log of multi-year (on average 12-year) average of parents' family income	0.40
	Log of multi-year (on average 8-year) average of family income			0.46
Solon (1992)	Log annual earnings in 1984	25–33	5-year average of father's log annual earnings	0.41
	Log hourly earnings in 1984		Father's log hourly earnings in 1967	0.29
	Log family income in 1984		Parents' log family income in 1967	0.48
	Log of ratio of family income to poverty line in 1984		Parents' log of ratio of family income to poverty line in 1967	0.48

[a] Couch and Dunn (1997) report a sample mean age of 24.9 in 1984.

Table 4
Least squares estimates of the elasticity of son's earnings with respect to parental income: NLS

Study	Son's outcome variable	Sons' age range	Measure of parental income	$\hat{\rho}$
Altonji and Dunn (1991)	Multi-year (up to 12-year) average of log annual earnings	24–39	Multi-year (up to 8-year) average of father's log annual earnings	0.18
	Multi-year (up to 12-year) average of log hourly wage		Multi-year (up to 8-year) average of father's log hourly wage	0.26
	Multi-year (up to 8-year) average of log family income		Multi-year (up to 7-year) average of father's log family income	0.27
			Multi-year (up to 7-year) average of mother's log family income	0.31
Couch and Lillard (1994)	Log annual earnings in 1980	28–38	Log of 4-year average of father's annual earnings	0.37
Peters (1992)	Multi-year (up to 4-year) average of log annual earnings	24–39	Multi-year (up to 5-year) average of father's log annual earnings	0.14
	Multi-year (up to 4-year) average of log family income		Multi-year (up to 5-year) average of parents' log family income	0.24
Zimmerman (1992)	Log annual earnings in 1981	29–39	4-year average of father's log annual earnings	0.54
	Log hourly wage in 1981		4-year average of father's log hourly wage	0.39

component of log earnings, 0.42 in the permanent component of log hourly wage, and 0.36 in log family income, as well as a mother-son correlation of 0.56 in log family income. Assuming instead that the transitory component follows a first-order autoregressive process, Zimmerman (1992) performs instrumental-variables estimation of Eq. (18) with the father's log single-year earnings X_{is} as an error-ridden proxy for the father's permanent status X_i and with X_{is} instrumented by the lead of its own quasi-first-difference. The central tendency of the resulting estimates is about 0.4.

Several other studies try instrumenting father's log single-year earnings with other characteristics of the father. Solon (1992) uses father's education as an instrument, Zimmerman (1992) uses the Duncan index of the prestige of the father's occupation, and Mulligan (1997) uses a variety of characteristics associated with the father's race, education, occupation, industry, and county of residence.[15] The resulting estimates tend to be larger than the least squares estimates, but, as shown in the appendix of Solon (1992), the instrumental-variables estimates are inconsistent for the intergenerational elasticity ρ if, conditional on father's permanent earnings, the instruments still would have independent predictive power for son's earnings. In that case, entering the instruments in the second stage of two-stage-least squares *only* as predictors of father's log single-year earnings induces a sort of omitted-variables inconsistency.

Section 3 concluded that the correlation among US brothers in their long-run earnings is about 0.4 or a bit higher. All in all, 0.4 or a bit higher also seems a reasonable guess of the intergenerational elasticity in long-run earnings for men in the United States. This figure is more than twice as high as most of the early estimates surveyed by Becker and Tomes, and it indicates that the United States is not nearly as close to the perfect mobility of society B as we used to think. Even so, an intergenerational elasticity of 0.4 for US men does not nearly account for all of the 0.4 brother correlation. As shown in Eq. (21), the brother correlation can be decomposed into the squared intergenerational elasticity plus a second component due to factors uncorrelated with parental income. Squaring a 0.4 intergenerational elasticity accounts for only 0.16 out of the 0.4 brother correlation. This suggests that, of the 40% or so of permanent earnings inequality that arises from the family and community background factors shared by brothers, probably only a minority share is related to parental income. At least as large a share seems to come from factors uncorrelated with parental income. What those mysterious factors are remains an important and challenging question for future research.

Although most of the new studies pertain to sons in the US longitudinal surveys, evidence is starting to accumulate for sons in some other countries. Table 5 summarizes the intergenerational elasticity estimates from Canada, Finland, Germany, Malaysia,

[15] In addition, studies that track the intergenerational progress of groups larger than families also can be reinterpreted as efforts to estimate Eq. (18) by instrumental variables. For example, Borjas' (1993, 1994) regressions of second-generation immigrants' status on the average status of the preceding generation of their ethnic group can be thought of as instrumental-variables estimation of Eq. (18) with ethnic-group dummy variables as the instruments.

Sweden, and the United Kingdom. Comparing estimates across countries is tricky because of the many differences across studies in their datasets and estimation methods. For example, Couch and Dunn's (1997) intergenerational elasticity estimate of 0.11 for Germany initially seems strikingly lower than most US estimates until one recalls that their parallel estimate for the United States (shown in Table 3) is only 0.13. Both of these small estimates probably are driven by Couch and Dunn's unusually young samples. Indeed, using data from more recent waves of the same German longitudinal survey, Wiegand (1997) obtains much larger intergenerational elasticity estimates by observing the sons at more mature ages.

Bjorklund and Jantti's (1997) study of Sweden also facilitates international comparisons by providing parallel estimates for the United States. Because their Swedish dataset includes father's education and occupation, but not father's earnings, Bjorklund and Jantti are unable to perform direct estimation of the regression of son's log earnings on father's log earnings. Instead, they first use data from the fathers' generation to estimate the regression of log earnings on education and occupation, use the results to construct a prediction of father's log earnings, and then estimate a second-stage regression of son's log earnings on father's predicted log earnings. This procedure is similar to the instrumental-variables estimation used in some US studies and is subject to the same inconsistency. Indeed, Bjorklund and Jantti find that, when they apply their two-stage procedure to PSID data from the United States, the resulting intergenerational elasticity estimate is greater than the estimate they obtain for the PSID when they estimate the direct regression of son's log earnings on father's actual log earnings. Since their two-stage estimate for Sweden is smaller than both the direct and two-stage estimates from the PSID, they suspect that intergenerational earnings mobility is greater in Sweden than in the United States. That conjecture is supported by Gustafsson's (1994) study, which estimates only a 0.14 intergenerational elasticity with a Swedish dataset that does contain data on father's income. As Gustafsson points out, his estimate is biased downward by his reliance on a single-year measure of father's income, but even a generous upward correction still produces an estimate considerably lower than most US estimates. The studies by Corak and Heisz (1998) and Jantti and Osterbacka (1996) suggest that Canada and Finland, like Sweden, are more mobile societies than the United States.

In contrast, the intergenerational elasticity estimates for the United Kingdom are quite high. Atkinson et al. (1983) estimate a 0.42 intergenerational elasticity even though their estimate is biased downward by reliance on a short-run measure of father's earnings and a homogeneous sample of fathers living in York's working-class neighborhoods. Dearden et al. (1997) estimate an even higher elasticity of 0.57, but their estimate may be biased upward by their prediction of father's earnings on the basis of his education and social class.

It is sometimes conjectured that intergenerational transmission of economic status is particularly strong in less developed countries.[16] Unfortunately, Lillard and Kilburn's (1995) study of Malaysia is the only one so far to use income data for both generations in a less developed country. Lillard and Kilburn conclude that, in Malaysia, "The earnings

Table 5
Estimates of the elasticity of son's earnings with respect to parental income in countries other than the United States

Study	Sample	Son's outcome variable	Sons' age range	Measure of parental income	$\hat{\rho}$
Atkinson et al. (1983)	Fathers in working-class neighborhoods of York, England, in 1950 and their sons	Log hourly earnings at survey date (1975–78)	Not reported[a]	Father's log weekly earnings in 1950	0.42
Bjorklund and Jantti (1997)	Swedish Level of Living Surveys	Log annual earnings in 1990	29–38	Prediction of father's log annual earnings based on his education and occupation	0.28
		Log individual income in 1990		Prediction of father's log individual income	0.36
Corak and Heisz (1998)	Canadian income tax records	Log individual income in 1994	28–31	Log of 5-year average of father's individual income	0.17
Couch and Dunn (1997)	German Socio-Economic Panel	Log of multi-year (up to 6-year) average of annual earnings	Not reported[b]	Log of multi-year (up to 6-year) average of father's annual earnings	0.11
Dearden et al. (1997)	British National Child Development Survey	Log weekly earnings in 1991	33	Prediction of father's log weekly earnings based on his education and social class	0.57
Gustafsson (1994)	Fathers in Stockholm, Sweden, in 1955 and their sons born in 1939–1946	4-year average of log individual income	31–41	Father's log individual income in 1955	0.14
Jantti and Osterbacka (1996)	Finnish censuses	Log annual earnings in 1990	30–40	2-year average of father's log annual earnings	0.22
Lillard and Kilburn (1995)	Malaysian Family Life Surveys	Log annual earnings in 1988	Not reported[c]	Father's log annual earnings in 1976–1977	0.26
Wiegand (1997)	German Socio-Economic Panel	Log monthly earnings in 1994	27–33	5-year average of father's log monthly earnings	0.34

[a] Atkinson et al. (1983) do not report an age range for their regression sample, but their Table 4.4 for a broader sample shows a very wide range, from under 25 to over 65.

[b] Couch and Dunn (1997) report a sample mean age of 22.8 in 1984.

[c] Lillard and Kilburn (1995) require their sons to be over 18 and report a sample mean age of 25.

link for sons is similar to that in the US..." It remains to be seen whether that finding will be replicated in other less developed countries.

It would be premature to reach firm conclusions on the basis of the available fragments of evidence, but so far the United Kingdom and United States do appear to be less mobile societies than Canada, Finland, and Sweden. A more thorough comparison across countries, preferably including less developed countries, may eventually prove to be a useful way of generating clues about the determinants of intergenerational transmission of earnings status.

All of the evidence discussed so far has been about intergenerational mobility for sons. Much less evidence is available on daughters' mobility. Presumably, this neglect of daughters (of which I have been as guilty as anyone) stems partly from our profession's usual sexism and partly from a recognition that, in societies in which married women's labor force participation rates are lower than men's, women's earnings may often be an unreliable indicator of their economic status. The latter, though, is no excuse for failing to analyze broader measures of daughters' adult status, such as family income. Indeed, this is what has been done in some of the few studies of daughters, which are summarized in Table 6.

Section 3's review of estimated sibling correlations in earnings found that the estimated sister-sister correlations tend to be about as large as the estimated brother-brother correlations. Similarly, if one compares the estimated daughter elasticities in Table 6 to the corresponding son estimates from the same studies in Tables 3–5, intergenerational transmission of economic status seems to be about as strong for daughters as for sons. Some of the largest elasticity estimates in Table 6 appear where the daughter's status is measured by her family income or by her husband's earnings. Indeed, Atkinson et al. (1983) estimate that the elasticity of the daughter's husband's earnings with respect to her father's earnings is just as great as the elasticity of a son's earnings with respect to his own father's earnings. Similarly, the part of Altonji and Dunn's (1991) NLS study that tabulates sample correlations among multi-year averages of age-adjusted log earnings reports a correlation of 0.26 between husbands and their fathers-in-law, as compared to a correlation of 0.22 between sons and fathers. These results suggest that, as new studies of daughters' mobility are undertaken to redress the gender imbalance in this area of research, serious attention should be given to the role of assortative mating in intergenerational transmission of economic status.[17]

For the convenience of expressing intergenerational mobility in terms of a single parameter, most of the evidence discussed in this section has involved linear regressions of son's or daughter's log income variables on parents' log income variables. The implicit assumption of a constant-elasticity relationship between child's and parents' incomes,

[16] For example, Lam and Schoeni (1993) state as a "stylized fact" that "intergenerational mobility is lower in developing countries, with family background playing a more important role in determining earnings," but they describe the supporting evidence as "impressionistic."

[17] In their study of Brazil, Lam and Schoeni (1993) stress the importance of assortative mating in explaining the strong association between husband's wage rate and father-in-law's years of schooling.

Table 6
Intergenerational elasticity estimates for daughters

Study	Sample	Daughter's outcome variable	Daughters' age range	Measure of parental income	$\hat{\rho}$
Altonji and Dunn (1991)	NLS	Multi-year (up to 11-year) average of log annual earnings	24–38	Multi-year (up to 8-year) average of father's log annual earnings	0.22
		Multi-year (up to 11-year) average of log hourly wage		Multi-year (up to 8-year) average of father's log hourly wage	0.23
		Multi-year (up to 8-year) average of log family income		Multi-year (up to 7-year) average of father's log family income	0.26
				Multi-year (up to 7-year) average of mother's log family income	0.37
Atkinson et al. (1983)	Fathers in working-class neighborhoods of York, England, in 1950 and their daughters	Husband's log hourly earnings at survey date (1975–78)	Not reported[a]	Father's log weekly earnings in 1950	0.45
Dearden et al. (1997)	British National Child Development Survey	Log weekly earnings in 1991	33	Prediction of father's log weekly earnings based on his education and social class	0.68
Minicozzi (1997)	PSID	Log of 2-year average of annual earnings	28–29	Log of estimated present discounted value of parents' lifetime earnings	0.41
Peters (1992)	NLS	Multi-year (up to 4-year) average of log annual earnings	23–38	Multi-year (up to 5-year) average of father's log annual earnings	0.11
		Multi-year (up to 4-year) average of log family income		Multi-year (up to 5-year) average of parents' log family income	0.28
Shea (1997)	PSID	Log of multi-year (on average 8-year) average of annual earnings	25–40	Log of multi-year (on average 12-year) average of parents' family income	0.54
		Log of multi-year (on average 8-year) average of family income			0.39

[a] Atkinson et al. (1983) do not report an age range for their regression sample, but their Table 4.4 for a broader sample shows a very wide range, from under 25 to over 65.

however, must surely be at least somewhat false, and a few studies have begun to investigate the particulars of how it is false. Behrman and Taubman (1990), Solon (1992), and Corak and Heisz (1998) experiment with estimating more flexibly specified regression functions and find some evidence that intergenerational regression to the mean is stronger from the bottom of the earnings distribution than from the top. Atkinson et al. (1983), Zimmerman (1992), and Dearden et al. (1997) report a similar pattern in their estimated transition matrices. Finally, foreshadowing the topic of Section 5, Minicozzi (1997) presents figures suggestive of an interaction effect between parental income and average neighborhood income.

4.3. What we have learned and what we still do not know

Most of the evidence from the many recent empirical studies of intergenerational mobility indicates that intergenerational earnings elasticities are substantial and are larger than we used to think. Even so, comparing these new estimates to the estimated sibling correlations reviewed in Section 3 suggests that much (and probably most) of what matters about which family one comes from is uncorrelated with parental earnings or income. The question of what those uncorrelated factors are and why they matter so much is an important and formidable challenge for future research.

Learning that intergenerational earnings elasticities are larger than we used to think is a real step forward, but, as is so often the case in scholarly research, improving our answer to one question leads immediately to harder questions. Now that we know parental income is a fairly strong predictor of offspring's earnings, it becomes that much more important to find out which of the causal processes modeled theoretically in Section 2 are mainly responsible for the empirically observed intergenerational associations of earnings. Are earnings correlated across generations because high-income parents have the wherewithal to invest more in their children's human capital, or because the genetic or cultural traits that contributed to the parents' high earnings are passed on to the children? Some recent studies, such as Mayer (1997) and Shea (1997), attempt to sort out the sources of intergenerational earnings transmission, but their identifying assumptions are not that compelling.[18] Finding more credible empirical leverage for answering this very difficult question will require extraordinary ingenuity.

[18] For example, one of Mayer's approaches is to compare the predictive power of parental income received when the child is ages 13–17 to the predictive power of parental income received after the grown child's earnings are observed. She identifies the separate causal processes by assuming that the predictive ability of parental income received after the child grows up cannot reflect investment of that income in the child's human capital. As Shea points out, however, parental investment while the child is still at home may be influenced by the parents' anticipation of future income, and furthermore their measured income after the child grows up may serve as a proxy for imperfectly measured income during the child's youth. Shea uses a different approach for estimating the extent to which parental income matters because higher parental income enables greater parental investment in the child's human capital. His estimates, however, are imprecise and are based on the dubious assumption that parental income variation predicted by the parents' union status, industry affiliation, and job displacement experience is uncorrelated with the genetic and cultural endowments that parents pass on to their children.

5. Neighborhood effects[19]

As emphasized in Section 3, the sibling correlation in earnings is generated by some combination of the family and community origins shared by siblings. Traditionally, the research literature has concentrated mainly on the influence of family origins, but some recent research – motivated initially by concerns about the impact of "underclass" neighborhoods on the children that grow up in them[20] – has turned to the influence of community origins. As detailed in Jencks and Mayer (1990), neighborhoods may influence children in numerous ways: through peer influences, through role-modeling and enforcement of social norms by adult residents of the community, and through influences of neighborhood institutions (including effects of school quality). Several recent theoretical analyses have modeled the contribution of such neighborhood influences to inequality, intergenerational mobility, and economic growth.[21]

The first part of this section will again extend the statistical model developed in the previous sections. This extended model will clarify the conceptual connections among sibling correlations, neighbor correlations, and regression studies of neighborhood effects. The second part will briefly review the empirical literature on neighborhood effects, and the third part will summarize what has been learned so far and what remains to be studied.

5.1. Statistical model

Two subscripts – j for the sibling and i for the family – are no longer enough. Now we need a third subscript h for the neighborhood (h for "hood?"). Rewrite Eq. (11) from Section 3 as

$$y_{hij} = a_{hi} + b_{hij} \tag{25}$$

where y_{hij} is some measure of the adult socioeconomic status of sibling j from family i in neighborhood h and, as before, a_{hi} is the component common to siblings from that family and b_{hij} is an orthogonal component idiosyncratic to the jth sibling. Next, decompose a_{hi} as

$$a_{hi} = \gamma' F_{hi} + \delta' N_h, \tag{26}$$

where F_{hi} is the vector containing all the family background characteristics (including but not restricted to parental income X_{hi}) that influence y_{hij} and N_h is the vector containing all the neighborhood background characteristics that influence y_{hij}. The terms $\gamma' F_{hi}$ and $\delta' N_h$ probably are positively correlated because advantaged families sort into advantaged neighborhoods.

Substituting Eq. (26) into Eq. (25) yields

$$y_{hij} = \gamma' F_{hi} + \delta' N_h + b_{hij}, \tag{27}$$

[19] This section draws heavily from Solon et al. (1997).

[20] See, for example, Murray (1984) and Wilson (1987).

[21] See, for example, Benabou (1996a,b), Durlauf (1996), and Kremer (1997).

which expresses the individual's adult status as a regression function of family background characteristics F_{hi}, the characteristics N_h of the neighborhood(s) the individual grew up in, and an error term b_{hij}. As will be discussed in the next subsection, Eq. (27) is the equation that regression studies of neighborhood effects heroically attempt to estimate. The attempt is heroic because we researchers are not omniscient enough to observe all elements of F_{hi} and N_h, and because even the elements we do observe often are measured with error.

Given the model's assumptions, the variance of y_{hij} is

$$\sigma_y^2 = \mathrm{Var}(\gamma' F_{hi}) + \mathrm{Var}(\delta' N_h) + 2\mathrm{Cov}(\gamma' F_{hi}, \delta' N_h) + \mathrm{Var}(b_{hij}), \tag{28}$$

and the covariance in y_{hij} between siblings j and j' from the same family is

$$\mathrm{Cov}(y_{hij}, y_{hij'}) = \mathrm{Var}(\gamma' F_{hi}) + \mathrm{Var}(\delta' N_n) + 2\mathrm{Cov}(\gamma' F_{hi}, \delta' N_h). \tag{29}$$

Eq. (29) formalizes the obvious point that siblings have correlated outcomes because they share both family and community origins. Sibling correlations alone cannot identify the separate effects of family and neighborhood origins, but additional information might be gleaned from the covariance between neighboring children from different families in the same neighborhood h:

$$\mathrm{Cov}(y_{hij}, y_{hi'j'}) = \mathrm{Cov}(\gamma' F_{hi}, \gamma' F_{hi'}) + \mathrm{Var}(\delta' N_h) + 2\mathrm{Cov}(\gamma' F_{hi}, \delta' N_h). \tag{30}$$

Eq. (30) formalizes another obvious point – that neighbors have correlated outcomes not only because they share community origins, but also because their family backgrounds are somewhat similar. The neighbor covariance in Eq. (30) is smaller than the sibling covariance in Eq. (29) because the neighboring children's families are merely somewhat similar, not identical. The second term in Eq. (30) unambiguously reflects neighborhood effects, but the first term unambiguously stems from family effects, and the attribution of the third term between family and neighborhood effects is inherently ambiguous. Therefore, one can view an estimate of the neighbor covariance as setting an estimated upper bound on the portion of the population variance in y_{hij} that is due to variation in neighborhood background. Furthermore, as explained in Altonji (1988) and Solon et al. (1997), that upper bound can be tightened by regression adjustments that partial out part of the first term in Eq. (30).[22]

5.2. Empirical studies

Most of the empirical studies of neighborhood effects have estimated regression equations of the form of Eq. (27).[23] Examples from the US literature include Brooks-Gunn et al.

[22] Such regression adjustments, however, may partial out indirect neighborhood effects that operate through their influence on parental characteristics. Suppose, for example, that living in a better neighborhood enables the parents to obtain higher-paying jobs and that the parents' increased income benefits the children's later socioeconomic outcomes. Controlling for parental income would subtract out this indirect neighborhood effect. Regression-adjusted estimates of neighbor correlations therefore should be viewed as bounding the *direct* effects of neighborhoods on children's outcomes.

(1993), Case and Katz (1991), Clark (1992), Corcoran et al. (1992), Crane (1991), Datcher (1982), Duncan (1994), and Kremer (1997). Only a few have examined neighborhood effects on children's later earnings; most have considered other outcomes, most frequently educational attainment. The studies also differ with respect to which family background variables they control for and which neighborhood characteristics they include in the N_h vector. Of course, none achieves the omniscience required to account for all relevant elements of N_h and F_{hi} and to measure them all accurately.

The results of the studies have been mixed, but it seems fair to say that, once a relatively thorough set of family background characteristics is controlled for, it is surprisingly difficult to produce robust evidence of strong neighborhood effects. Corcoran et al. (1992), for example, estimate small coefficients for their neighborhood variables and find that "F-tests of the joint hypothesis that all five community variables... have zero coefficients accept the hypothesis at the 0.05 level in the equations for son's earnings and income." Even studies that report larger estimated neighborhood effects sometimes exhibit symptoms that the results are fragile. Crane (1991), for example, candidly acknowledges that he settled on percentage of workers with professional or managerial jobs as the key neighborhood variable in his analysis of 1970 census data only after trying and discarding fifteen other neighborhood characteristics. Given the well-known dangers of "data mining," it should not be surprising that Clark's (1992) replication of Crane's study with 1980 census data fails to reproduce his pattern of results.

The fragility of existing estimates of neighborhood effects can be interpreted in more than one way. One possibility is simply that neighborhoods really do not matter that much. Indeed, Jencks and Mayer (1990), Corcoran et al. (1992), and Evans et al. (1992) all have suggested the possibility that some estimates of neighborhood effects are as large as they are partly because the neighborhood variables are serving as proxies for unmeasured aspects of family background. For example, Borjas (1995) uses data from the National Longitudinal Survey of Youth to estimate regressions of son's education or log wage on father's economic status and the average economic status of the family's ethnic group in the father's generation. He finds that his coefficient estimates become smaller when he controls for a vector of neighborhood dummy variables. The apparent importance of the neighborhood variables leads him to infer that "neighborhood characteristics influence intergenerational mobility." His regressions, however, include only one measure of parental status – either the father's years of education or the average log wage in the father's occupation. With such sparse controls for parental status, it is quite possible that the estimated neighborhood coefficients largely reflect the effects of unmeasured aspects of family background. And, because no study can possibly control for all imaginable aspects of family background, it is inevitable that conventional regression analyses will remain susceptible to this problem.[24]

Another possibility is that neighborhoods matter a lot, but their effects are hard to detect with the methods that have been used. Perhaps researchers have defined neighborhoods

[23] Some of the studies have estimated nonlinear regression models for binary outcome variables.

inappropriately or have focused on unimportant neighborhood characteristics while overlooking important ones. This point is reminiscent of the finding, emphasized in Section 3, that sibling correlations in socioeconomic status far exceed what has been explained by any particular *measured* aspects of the siblings' shared background. Similarly, Hanushek's (1986) survey of the related literature on school effects suggests that schools do matter even though it is difficult to attribute their effects to any particular commonly measured characteristics of schools.

Serious attention to this possibility is encouraged by an unusual study by Rosenbaum (1991). This study is based on Chicago's Gautreaux program, which relocated black residents of public housing to subsidized private apartments in Chicago and its suburbs. Rosenbaum claims that the process by which program applicants were allocated between city apartments and suburban apartments was essentially random, and then he compares outcomes between those that moved to the predominantly middle-class suburbs and those that moved within the city. The children from the families that moved to the suburbs were much less likely to drop out of high school and had considerably higher rates of college attendance, employment, and good pay. Although the study's sample is very small and rather special, the results reinforce the possibility that neighborhoods may exert important influences. Fischer (1991), however, reports replicating only some of Rosenbaum's Gautreaux results with a similar dataset from Cincinnati. Fortunately, the currently ongoing Moving to Opportunity study, funded by the US Department of Housing and Urban Development, is imitating the Gautreaux "experiment" on a larger scale in Baltimore, Boston, Chicago, Los Angeles, and New York City. Once the results of this social experiment become available, they may add substantially to our limited information on neighborhood influences.

Another approach to exploring the contribution of neighborhood background to inequality is to measure the neighbor correlation in y_{hij}, that is, the ratio of the neighbor covariance in Eq. (30) to the variance in Eq. (28). This approach sidesteps the questions of which neighborhood variables are the important ones and how they ought to be measured. Much as the sibling correlation indicates the proportion of the variance in y_{hij} due to disparities in family and neighborhood background variables (even unobserved ones), the neighbor correlation gives an upper bound on how much of the variance arises from neighborhood variables. It identifies only an upper bound because, as explained in the previous subsection, the neighbor covariance reflects the effects of neighbors' similar family background as well as the influence of their shared community background.

Solon et al. (1997) apply this approach to PSID data with years of education as the

[24] Some studies (Aaronson, 1995; Plotnick and Hoffman, 1995) have tried to control for family background by relating between-sibling differences in outcomes to between-sibling differences in neighborhood environment. It remains unclear, however, whether the observed between-sibling differences in outcomes are really caused by the neighborhood differences or by other factors associated with changing neighborhoods (e.g., divorce of the parents or a parent's job loss). The possibility that using between-sibling variation may aggravate rather than reduce endogeneity bias has been clearly discussed by Griliches (1979) and Card (1995) in a different context (estimating earnings returns to schooling).

outcome variable y_{hij}. They estimate the sibling correlation in years of education at a little more than 0.5, which is typical for US siblings data. In contrast, their estimate of the neighbor correlation is less than 0.2, even though this correlation still encompasses some effects of family background. The comparison suggests that the sibling resemblance in educational attainment is generated mostly by something about family background rather than neighborhood background. Furthermore, once a portion of the family effect is partialed out by regression adjustments for a few observable family characteristics, the estimated proportion of the variance in educational attainment that can be ascribed to neighborhood factors drops below 0.1. Jencks and Brown (1975) and Altonji (1988) report similar results for correlations between students in the same high school. It remains to be seen whether these results will persist when earnings are used as the outcome variable.

Even if they do, that will not deny that neighborhoods matter to some degree in determining earnings.[25] Nor will it deny that neighborhoods may exert quite large effects on some people. Even if neighborhoods cannot account for much of the population-wide variance in outcomes, children growing up in extreme neighborhood environments or with special sensitivity to those environments may be greatly influenced by their neighborhoods. The families in the Gautreaux program may be a prime example. These families began in extremely disadvantaged communities, and their choice to apply to the program presumably reflected a belief that changing neighborhoods would make a big difference in their lives. If anyone should exhibit large neighborhood effects, the Gautreaux families should.

5.3. What we have learned and what we still do not know

Numerous researchers have conducted regression studies of neighborhood effects, but these studies have been inconclusive and are likely to remain so. The ongoing Moving to Opportunity project and measurement of the correlation between neighboring children in their later earnings as adults are two promising alternatives for enhancing our very limited knowledge about the importance of community origins as a source of earnings inequality.

If it turns out that something about neighborhood background matters a great deal, we still will be left with the question of what that something is. Peer group effects? Role-modeling by adults in the community? Effects of neighborhood institutions? Coming up with feasible and convincing research designs for sorting out different avenues for neighborhood effects will be a formidable challenge.

[25] Of course, introspection by most readers will reveal that our location choices are motivated partly by a belief that the neighborhoods we choose do have at least some effect on our children's outcomes. In accordance with that casual empiricism, Black's (1996) econometric evidence indicates that home purchasers do pay a premium for houses in the attendance districts of schools in which the students achieve higher average test scores.

6. Conclusions

Over the last decade or so, we have made considerable progress in measuring the intergenerational association in earnings and the overall impact of family and community origins on earnings. Newly available intergenerational data from longitudinal surveys of national probability samples have revealed that intergenerational influences are stronger than social researchers had believed in the 1970s and 1980s. For men in the United States, for example, it appears that the intergenerational earnings elasticity is somewhere around 0.4, which is twice what used to be viewed as an upper bound. The elasticity estimates for Canada, Finland, and Sweden are smaller than for the United States, but still are larger than would have been surmised from earlier research methodologies. In terms of the example from this chapter's introduction, we have learned that the societies we live in are more like society A and less like society B than we used to think.

In light of this new learning, concerns about inequality of opportunity no longer can be summarily dismissed on the ground that our societies nearly attain the perfect intergenerational mobility of society B. But whether our societies have *too much* inequality of opportunity (or not enough) remains quite open to debate. One's views should depend, among other things, on one's beliefs about *why* intergenerational influences on earnings are as strong (and as weak) as they are. Unfortunately, we remain fairly ignorant about the causal processes underlying the intergenerational transmission of earnings.[26] For example, we presently have very little empirical basis for assessing why parental income matters as much as it does. Is it because high-income parents are able to invest more in their children's human capital, or because the genetic or cultural traits that contributed to the parents' high earnings are passed on to the children? In any case, a comparison of sibling and intergenerational correlations suggests that much, perhaps most, of the intergenerational influence on earnings is unrelated to parental income. Where it does come from remains a fascinating and important puzzle for future research.

Further advances will not come easily. Undoubtedly, we economists will continue our usual estimation of regression models with the usual survey data, but, as discussed in Section 5, such analyses by themselves probably will not settle matters. We need to be alert for new, and sometimes rather peculiar, data that enable new perspectives. Some examples mentioned along the way in this chapter have been data on identical twins reared apart, adopted siblings not biologically related, the Gautreaux program, and the Moving to Opportunity project. International comparisons also may be illuminating. Of course, none of these natural and not-so-natural experiments achieves the ideal of a clean and definitive experiment, but judicious interpretation of the evidence they provide still may yield important clues about the processes underlying intergenerational influences on labor market status.

[26] Even when we have accumulated better evidence on the sources of intergenerational transmission, well informed, well-intentioned people still will differ in their policy views because of different value judgments about what constitutes a fair earnings distribution and about the extent to which efficiency losses should be suffered to achieve it.

References

Aaronson, Daniel (1995), "Using sibling data to estimate the impact of neighborhoods on children's educational outcomes", Unpublished.

Altonji, Joseph G. (1988), "The effects of family background and school characteristics on education and labor market outcomes", Unpublished.

Altonji, Joseph G. and Thomas A. Dunn (1991), "Relationships among the family incomes and labor market outcomes of relatives", Research in Labor Economics 12: 269–310.

Ashenfelter, Orley and Alan Krueger (1994), "Estimates of the economic returns to schooling from a new sample of twins", American Economic Review 84: 1157–1173.

Ashenfelter, Orley and David J. Zimmerman (1997), "Estimates of the returns to schooling from sibling data: fathers, sons and brothers", Review of Economics and Statistics 79: 1–9.

Atkinson, A.B., A.K. Maynard and C.G. Trinder (1983), Parents and children: incomes in two generations (Heinemann, London).

Aughinbaugh, Alison (1996), "Intergenerational income mobility in the United States: comment", Unpublished.

Baker, Michael (1997), "Growth rate heterogeneity and the covariance structure of life-cycle earnings", Journal of Labor Economics 15: 537–579.

Baker, Michael and Gary Solon (1997), "Earnings dynamics and inequality among Canadian men, 1976–1992: evidence from longitudinal income tax records", Unpublished.

Becker, Gary S. (1988), "Family economics and macro behavior", American Economic Review 78: 1–13.

Becker, Gary S. and Nigel Tomes (1979), "An equilibrium theory of the distribution of income and intergenerational mobility", Journal of Political Economy 87: 1153–1189.

Becker, Gary S. and Nigel Tomes (1986), "Human capital and the rise and fall of families", Journal of Labor Economics 4: S1–S39.

Behrman, Jere R. (1997), "Intrahousehold distribution and the family", in: Mark R. Rosenzweig and Oded Stark, eds., Handbook of population and family economics, Vol. 1A (North-Holland, Amsterdam) pp. 125–187.

Behrman, Jere R. and Paul Taubman (1985), "Intergenerational earnings mobility in the united states: some estimates and a test of Becker's intergenerational endowments model", Review of Economics and Statistics 67: 144–151.

Behrman, Jere R. and Paul Taubman (1990), "The intergenerational correlation between children's adult earnings and their parents' income: results from the Michigan Panel Survey of Income Dynamics", Review of Income and Wealth 36: 115–127.

Behrman, Jere R., Paul Taubman and Terence Wales (1977), "Controlling for and measuring the effects of genetics and family environment in equations for schooling and labor market success", in: Paul Taubman, ed., Kinometrics: determinants of socioeconomic success within and between families (North-Holland, Amsterdam) pp. 35–96.

Benabou, Roland (1996a), "Equity and efficiency in human capital investment: the local connection", Review of Economic Studies 63: 237–264.

Benabou, Roland (1996b), "Heterogeneity, stratification and growth: macroeconomic implications of community structure and school finance", American Economic Review 86: 584–609.

Betts, Julian (1996), "Is there a link between school inputs and earnings? Fresh scrutiny of an old literature", in: Gary Burtless, ed., Does money matter? The effect of school resources on student achievement and adult success (Brookings, Washington, DC) pp. 141–191.

Bjorklund, Anders (1993), "A comparison between actual distributions of annual and lifetime income: Sweden 1951–89", Review of Income and Wealth 39: 377–386.

Bjorklund, Anders and Markus Jantti (1997), "Intergenerational income mobility in Sweden compared to the United States", American Economic Review 87: 1009–1018.

Black, Sandra E. (1996), "Do 'better' schools matter? Parents think so!" Unpublished.

Borjas, George J. (1993), "The intergenerational mobility of immigrants", Journal of Labor Economics 11: 113–135.

Borjas, George J. (1994), "Long-run convergence of ethnic skill differentials: the children and grandchildren of the Great Migration", Industrial and Labor Relations Review 47: 553–573.

Borjas, George J. (1995), "Ethnicity, neighborhoods and human-capital externalities", American Economic Review 85: 365–390.

Bouchard, Thomas J., Jr., David T. Lykken, Matthew McGue, Nancy L. Segal and Auke Tellegen (1990), "Sources of human psychological differences: the Minnesota study of twins reared apart", Science 250: 223–228.

Bound, John, Zvi Griliches and Bronwyn H. Hall (1986), "Wages, schooling and IQ of brothers and sisters: do the family factors differ?" International Economic Review 27: 77–105.

Bowles, Samuel (1972), "Schooling and inequality from generation to generation", Journal of Political Economy 80: S219–S251.

Brittain, John (1977), The inheritance of economic status (Brookings, Washington DC).

Brooks-Gunn, Jeanne, Greg J. Duncan, Pamela Kato Klebanov and Naomi Sealand (1993), "Do neighborhoods influence child and adolescent development?" American Journal of Sociology 99: 353–395.

Buron, Lawrence (1994), "A study of the magnitude and determinants of intergenerational earnings mobility", PhD dissertation (University of Wisconsin).

Card, David (1995), "Earnings, schooling and ability revisited", Research in Labor Economics 14: 23–48.

Card, David and Alan B. Krueger (1996), "School resources and student outcomes: an overview of the literature and new evidence from North and South Carolina", Journal of Economic Perspectives 10: 31–50.

Case, Anne C. and Lawrence F. Katz (1991), "The company you keep: the effects of family and neighborhood on disadvantaged youths", Working paper no. 3705 (NBER, Cambridge, MA).

Chamberlain, Gary and Zvi Griliches (1975), "Unobservables with a variance-components structure: ability, schooling and the economic success of brothers", International Economic Review 16: 422–449.

Clark, Rebecca L. (1992), "Neighborhood effects on dropping out of school among teenage boys", Unpublished.

Corak, Miles and Andrew Heisz (1998), "Unto the sons: the intergenerational income mobility of Canadian men", Research paper no. 113 (Analytical Studies Branch, Statistics Canada).

Corcoran, Mary and Linda P. Datcher (1981), "Intergenerational status transmission and the process of individual attainment", in: Martha S. Hill, Daniel H. Hill and James N. Morgan, eds., Five thousand American families: patterns of economic progress, Vol. IX (Institute for Social Research, University of Michigan, Ann Arbor, MI).

Corcoran, Mary and Christopher Jencks (1979), "The effects of family background", in: Christopher Jencks et al., eds., Who gets ahead? (Basic Books, New York) pp. 50–84.

Corcoran, Mary, Christopher Jencks and Michael Olneck (1976), "The effects of family background on earnings", American Economic Review 66: 430–435.

Corcoran, Mary, Roger Gordon, Deborah Laren and Gary Solon (1990), "Effects of family and community background on economic status", American Economic Review 80: 362–366.

Corcoran, Mary, Roger Gordon, Deborah Laren and Gary Solon (1992), "The association between men's economic status and their family and community origins", Journal of Human Resources 27: 575–601.

Couch, Kenneth A. and Thomas A. Dunn (1997), "Intergenerational correlations in labor market status: a comparison of the United States and Germany", Journal of Human Resources 32: 210–232.

Couch, Kenneth A. and Dean R. Lillard (1994), "Sample selection rules and the intergenerational correlation of earnings: a comment on Solon and Zimmerman", Unpublished.

Crane, Jonathan (1991), "The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing", American Journal of Sociology 96: 1226–1259.

Datcher, Linda P. (1982), "Effects of community and family background on achievement", Review of Economics and Statistics 64: 32–41.

Dearden, Lorraine, Stephen Machin and Howard Reed (1997), "Intergenerational mobility in Britain", Economic Journal 107: 47–66.

Duncan, Greg J. (1994), "Families and neighbors as sources of disadvantage in the schooling decisions of white and black adolescents", American Journal of Education 103: 20–53.

Durlauf, Steven N. (1996), "A theory of persistent income inequality", Journal of Economic Growth 1: 75–93.
Eide, Eric and Mark Showalter (1997), "Factors affecting the transmission of earnings across generations: a quantile regression approach", Unpublished.
Erikson, Robert and John H. Goldthorpe (1992), The constant flux: a study of class mobility in industrial societies (Clarendon, Oxford, UK).
Evans, William N., Wallace E. Oates and Robert M. Schwab (1992), "Measuring peer group effects: a study of teenage behavior", Journal of Political Economy 100: 966–991.
Feldman, Marcus W., Sarah P. Otto and Freddy B. Christiansen (1996), "Genes, culture and inequality", Unpublished.
Fischer, Paul B. (1991), "Is housing mobility an effective anti-poverty strategy? An examination of the Cincinnati experience", Unpublished.
Friedman, Milton (1957), A theory of the consumption function (Princeton University Press, Princeton, NJ).
Ganzeboom, Harry B.G., Donald J. Treiman and Wout C. Ultee (1991), "Comparative intergenerational stratification research: three generations and beyond", Annual Review of Sociology 17: 277–302.
Goldberger, Arthur S. (1979), "Heritability", Economica 46: 327–347.
Goldberger, Arthur S. (1989), "Economic and mechanical models of intergenerational transmission", American Economic Review 79: 504–513.
Gordon, Roger H. (1984), Differences in earnings and ability (Garland, New York).
Greene, William H. (1997), Econometric analysis, 3rd editon (Prentice Hall, Upper Saddle River).
Griliches, Zvi (1979), "Sibling models and data in economics: beginnings of a survey", Journal of Political Economy 87: S37–S64.
Gustafsson, Bjorn (1994), "The degree and pattern of income immobility in Sweden", Review of Income and Wealth 40: 67–86.
Haider, Steven J. (1997), "Earnings instability and earnings inequality of males in the United States: 1967–1991", Unpublished.
Hanushek, Eric A. (1986), "The economics of schooling: production and efficiency in public schools", Journal of Economic Literature 24: 1141–1177.
Hauser, Robert M. and William H. Sewell (1986), "Family effects in simple models of education, occupational status and earnings: findings from the Wisconsin and Kalamazoo studies", Journal of Labor Economics 4: S83–S115.
Herrnstein, Richard J. and Charles Murray (1994), The bell curve: intelligence and class structure in American life (Free Press, New York).
Isacsson, Gunnar (1997), "Estimates of the return to schooling in Sweden from a large sample of twins", Unpublished.
Jantti, Markus and Eva Osterbacka (1996), "How much of the variance in income can be attributed to family background? Empirical evidence from Finland", Unpublished.
Jencks, Christopher S. and Marsha D. Brown (1975), "Effects of high schools on their students", Harvard Educational Review 45: 273–324.
Jencks, Christopher and Susan E. Mayer (1990), "The social consequences of growing up in a poor neighborhood", in: Laurence Lynn, Jr. and Michael McGeary, eds., Inner-city poverty in the United States (National Academy Press, Washington, DC) pp. 111–186.
Jencks, Christopher, Marshall Smith, Henry Acland, Mary Jo Bane, David Cohen, Herbert Gintis, Barbara Heyns and Stephan Michelson (1972), Inequality: a reassessment of the effect of family and schooling in America (Basic Books, New York).
Kearl, J.R. and Clayne L. Pope (1986), "Unobservable family and individual contributions to the distributions of income and wealth", Journal of Labor Economics 4: S48–S79.
Krasker, William S., Edwin Kuh and Roy E. Welsch (1983), "Estimation for dirty data and flawed models", in: Zvi Griliches and Michael D. Intriligator, eds., Handbook of econometrics, Vol. 1 (North-Holland, Amsterdam) pp. 651–698.

Kremer, Michael (1997), "How much does sorting increase inequality?" Quarterly Journal of Economics 112: 115–139.

Lam, David and Robert F. Schoeni (1993), "Effects of family background on earnings and returns to schooling: evidence from Brazil", Journal of Political Economy 101: 710–740.

Lillard, Lee A. and M. Rebecca Kilburn (1995), "Intergenerational earnings links: sons and daughters", Unpublished.

Lillard, Lee A. and M. Rebecca Kilburn (1996), "Assortative mating and family links in permanent earnings", Unpublished.

Lillard, Lee A. and Robert T. Reville (1996), "Intergenerational mobility in earnings and occupational status", Unpublished.

Lillard, Lee A. and Robert J. Willis (1978), "Dynamic aspects of earning mobility", Econometrica 46: 985–1012.

Mayer, Susan E. (1997), What money can't buy: family income and children's life chances (Harvard University Press, Cambridge, MA).

Miller, Paul, Charles Mulvey and Nick Martin (1995), "What do twins studies reveal about the economic returns to education? A comparison of Australian and U.S. findings", American Economic Review 85: 586–599.

Minicozzi, Alexandra L. (1997), "Nonparametric analysis of intergenerational income mobility", PhD dissertation (University of Wisconsin).

Mulligan, Casey B. (1997), Parental priorities and economic inequality (University of Chicago Press, Chicago, IL).

Murray, Charles (1984), Losing ground: American social policy, 1950–1980 (Basic Books, New York).

Olneck, Michael R. (1977), "On the use of sibling data to estimate the effects of family background, cognitive skills and schooling: results from the Kalamazoo brothers study", in: Paul Taubman, ed., Kinometrics: determinants of socioeconomic success within and between families (North-Holland, Amsterdam).

Peters, H. Elizabeth (1992), "Patterns of intergenerational mobility in income and earnings", Review of Economics and Statistics 74: 456–466.

Plomin, Robert and Denise Daniels (1987), "Why are children in the same family so different from one another?" Behavioral and Brain Sciences 10: 1–16.

Plomin, Robert and Stephen A. Petrill (1997), "Genetics and intelligence: what's new?" Intelligence 24: 53–77.

Plotnick, Robert D. and Saul D. Hoffman (1995), "Fixed effect estimates of neighborhood effects", Working paper no. 95/06 (Department of Economics, University of Delaware).

Reville, Robert T. (1995), "Intertemporal and life cycle variation in measured intergenerational earnings mobility", Unpublished.

Rosenbaum, James E. (1991), "Black pioneers – do their moves to the suburbs increase opportunity for mothers and children?" Housing Policy Debate 2: 1179–1213.

Rouse, Cecilia Elena (1997), "Further estimates of the economic return to schooling from a new sample of twins", Working paper no. 388 (Industrial Relations Section, Princeton University, Princeton, NJ).

Shea, John (1997), "Does parents' money matter?" Unpublished.

Solon, Gary (1992), "Intergenerational income mobility in the United States", American Economic Review 82: 393–408.

Solon, Gary (1994), "Comments on 'Sample selection rules and the intergenerational correlation of earnings: a comment on Solon and Zimmerman' by Couch and Lillard", Unpublished.

Solon, Gary, Mary Corcoran, Roger Gordon and Deborah Laren (1988), "Sibling and intergenerational correlations in welfare program participation", Journal of Human Resources 23: 388–396.

Solon, Gary, Mary Corcoran, Roger Gordon and Deborah Laren (1991), "A longitudinal analysis of sibling correlations in economic status", Journal of Human Resources 26: 509–534.

Solon, Gary, Marianne E. Page and Greg J. Duncan (1997), "Correlations between neighboring children in their subsequent educational attainment", Unpublished.

Taubman, Paul (1976), "The determinants of earnings: genetics, family and other environments; a study of white male twins", American Economic Review 66: 858–870.

Weiss, Yoram (1997), "The formation and dissolution of families: why marry? Who marries whom? And what

happens upon divorce", in: Mark R. Rosenzweig and Oded Stark, eds., Handbook of population and family economics, Vol. 1A (North-Holland, Amsterdam) pp. 81–123.

Wiegand, Johannes (1997), "Intergenerational earnings mobility in Germany", Unpublished.

Willis, Robert J. (1986), "Wage determinants: a survey and reinterpretation of human capital earnings", in: Orley C. Ashenfelter and Richard Layard, eds., Handbook of labor economics, Vol. 1 (North-Holland, Amsterdam) pp. 525–602.

Wilson, William Julius (1987), The truly disadvantaged: the inner city, the underclass and public policy (University of Chicago Press, Chicago IL).

Zimmerman, David J. (1992), "Regression toward mediocrity in economic stature", American Economic Review 82: 409–429.

Chapter 30

THE CAUSAL EFFECT OF EDUCATION ON EARNINGS

DAVID CARD*

Department of Economics, University of California at Berkeley

Contents

Abstract 1802
JEL codes 1802
1 Introduction and overview 1802
2 The human capital earnings function 1803
2.1 Functional form 1804
2.2 Measurement of education 1806
2.3 Which measure of earnings? 1808
2.4 Summary 1809
3 Causal modelling of the return to education 1810
3.1 Theoretical issues 1810
3.2 Observed schooling and earnings outcomes 1813
3.3 Measurement error 1815
3.4 Instrumental variables estimates of the return to schooling 1817
3.5 Limitations of instrumental variables 1819
3.6 Family background 1822
3.7 Models for siblings and twins 1826
3.8 Summary 1831
4 A selective review of recent empirical studies 1834
4.1 Instrumental variables based on institutional features of the school system 1834
4.2 Estimators using family background as a control or instrument 1842
4.3 Studies of education and earnings using twins 1846
4.4 Direct evidence on the heterogeneity in returns to education 1852
5 Conclusions 1855
Appendix A 1856
A.1 OLS estimation of a random coefficients model 1856
A.2 Estimation of a random coefficients model 1857
A.3 Measurement error in a bivariate regression model 1858
References 1859

* I am grateful to David Lee and Gena Estes for research assistance, to Orley Ashenfelter, Alan Krueger, and James Powell for helpful discussions, and to Michael Boozer, Ken Chay, Andrew Hildreth and Gary Solon for comments on earlier drafts that substantially improved the chapter. This research was funded in part by a grant from the NICHD.

Handbook of Labor Economics, Volume 3, Edited by O. Ashenfelter and D. Card

Abstract

This paper surveys the recent literature on the causal relationship between education and earnings. I focus on four areas of work: theoretical and econometric advances in modelling the causal effect of education in the presence of heterogeneous returns to schooling; recent studies that use institutional aspects of the education system to form instrumental variables estimates of the return to schooling; recent studies of the earnings and schooling of twins; and recent attempts to explicitly model sources of heterogeneity in the returns to education. Consistent with earlier surveys of the literature, I conclude that the average (or average marginal) return to education is not much below the estimate that emerges from a standard human capital earnings function fit by OLS. Evidence from the latest studies of identical twins suggests a small upward "ability" bias – on the order of 10%. A consistent finding among studies using instrumental variables based on institutional changes in the education system is that the estimated returns to schooling are 20–40% above the corresponding OLS estimates. Part of the explanation for this finding may be that marginal returns to schooling for certain subgroups – particularly relatively disadvantaged groups with low education outcomes – are higher than the average marginal returns to education in the population as a whole.

JEL codes: I20; J30

1. Introduction and overview

Education plays a central role in modern labor markets. Hundreds of studies in many different countries and time periods have confirmed that better-educated individuals earn higher wages, experience less unemployment, and work in more prestigious occupations than their less-educated counterparts.[1] Despite the overwhelming evidence of a positive correlation between education and labor market status, social scientists have been cautious to draw strong inferences about the causal effect of schooling. In the absence of experimental evidence, it is very difficult to know whether the higher earnings observed for better-educated workers are *caused* by their higher education, or whether individuals with greater earning capacity have chosen to acquire more schooling.

Economists' interest in this issue was stimulated in the late 1950s by growth accounting exercises which found that rising education levels could explain much of post-war US productivity growth, leaving little room for technological change (see, e.g., Becker, 1964; Griliches, 1970). Skeptics noted that this conclusion was only valid if the observed cross-sectional earnings differences between education groups reflected true productivity differentials, rather than inherent ability differences that happened to be correlated with education (e.g., Denison, 1964). The emergence of large-scale microeconomic datasets in the 1960s lead to an outpouring of research on education and earnings, much of it focussed on the issue of "ability bias" in the earnings differentials between more- and less-educated

[1] See Cohn and Addison (1997) for a selective review of recent international studies, and Psacharopoulos (1985, 1994) for a broad overview of the international literature on schooling and earnings.

workers. In his landmark survey of the 1960s and 1970s literature, Griliches (1977) concluded that such biases were small – potentially even smaller than other biases that lead measured earnings differences to *understate* the causal effect of education. In his earlier review of the evidence, Becker (1964) had similarly concluded that ability biases were overstated by critics of the human capital paradigm.[2] Despite the careful reasoning of these earlier surveys, however, many analysts continue to believe that the measured partial correlation between schooling and earnings significantly overstates the true causal effect of education, and that findings to the contrary are counter-intuitive.

The aim of this chapter is to survey and interpret some of the most recent evidence on the causal relationship between schooling and earnings. I focus on four key areas of research:

1. theoretical and econometric advances in modelling the causal effect of education in the presence of heterogenous returns to schooling;
2. recent studies that use institutional aspects of the education system as "exogenous" sources of variation in education outcomes;
3. recent studies of the earnings and schooling outcomes of twins;
4. recent studies that explicitly model heterogeneity in the returns to education across groups or individuals.

A unifying theme in much of this work is that the return to education is not a single parameter in the population, but rather a random variable that may vary with other characteristics of individuals, such as family background, ability, or level of schooling. In my opinion, this broader view of the effect of education helps to reconcile the various findings in the literature, and provides a useful framework for generating new hypotheses and insights about the connection between education and earnings.

The chapter begins with a brief overview of the so-called human capital earnings function, which is the primary econometric model that economists use to measure the return to education. I then present an extended discussion of a simple theoretical model of endogenous schooling that is helpful in interpreting recent empirical studies. Finally, I present a selective review and synthesis of some of the most interesting new work on education and earnings.

2. The human capital earnings function

Recent studies of education and wage determination are almost always embedded in the framework of Mincer's (1974) human capital earnings function (HCEF). According to this model, the log of individual earnings (y) in a given time period can be decomposed into an

[2] Becker (1964, p. 88, footnote 30) offered the following interpretation of the prevailing opinion on the importance of ability biases: "A more cynical explanation would be that vocal observers are themselves primarily successful college graduates and, therefore, naturally biased toward the view that ability is a major cause of the high earnings received by college graduates."

additive function of a linear education term and a quadratic experience term:

$$\log y = a + bS + cX + dX^2 + e, \tag{1}$$

where S represents years of completed education, X represents the number of years an individual has worked since completing schooling, and e is a statistical residual. In the absence of direct information on experience Mincer proposed the use of "potential experience": the number of years an individual of age A could have worked, assuming he started school at age 6, finished S years of schooling in exactly S years, and began working immediately thereafter: $X = A - S - 6$. Although Mincer derived this equation from a theoretical model of schooling choice and post-schooling training decisions, the basic patterns of variation of earnings by age and education had been known at least since the early 1950s (e.g., Miller, 1955).[3] Thus the HCEF can be seen as an extraordinarily successful marriage of inductive and deductive reasoning.

2.1. Functional form

The simple specification of Eq. (1) immediately raises a number of questions that have been addressed directly and indirectly over the past 20 years. Many of these concern functional form. Mincer's equation can be regarded as an approximation to a general functional form,

$$\log y = F(S,A) + e.$$

Since both S and A are measured as discrete variables in most datasets, the function $F(\,)$ can be estimated non-parametrically by including a complete set of dummy variables for all (S,A) pairs, or by using non-parametric smoothing methods (e.g., kernel density estimators) in smaller datasets.[4] Alternatively, researchers have added higher-order terms in schooling and age or experience to (1) and examined the improvement in fit relative to Mincer's original specification. A comprehensive study along the latter lines by Murphy and Welch (1990) concluded that a generalization of Mincer's model

$$\log y = a + bS + g(X) + e, \tag{1'}$$

where g is a third or possibly fourth-order polynomial, provides a significant improvement in fit.

Some recent evidence on the shape of the $F(\,)$ function and the performance of a specification like (1′) is provided in Fig. 1, which shows actual age-earnings profiles for

[3] Miller (1955, pp. 64–67) displays the age profiles of annual earnings data for men in the 1950 Census for three different education groups and remarks on both the concave nature of these profiles, and the fact that the profile for better-educated men peaks about 10 years later than the profile for less-educated men. Miller's analysis of the 1960 Census data (Miller, 1966) confirmed these same tendencies.

[4] In most US datasets, for example, S takes on 18 or 20 discrete values and A ranges from 16 to 66, implying a maximum of about 1000 points in the range of $F(\,)$. Zheng (1996) uses formal testing methods to compare the fit of expanded various versions of (1) to kernel density estimates using March 1990 Current Population Survey data.

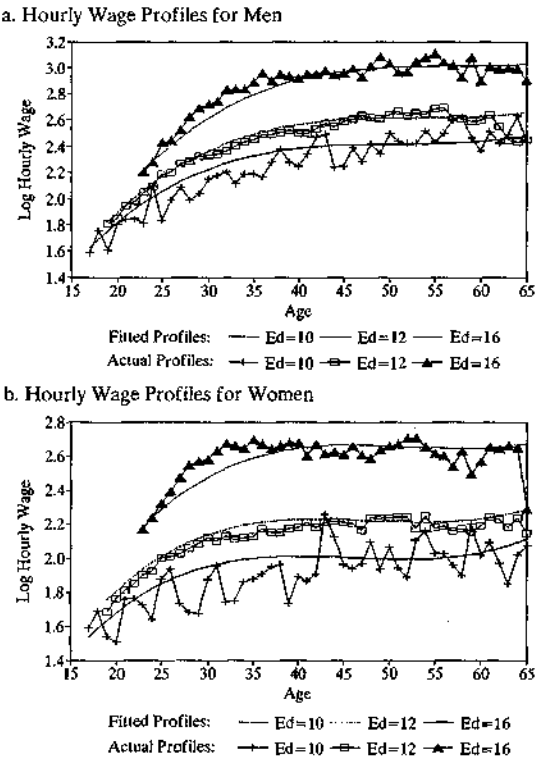

Fig. 1. Age profiles of hourly wags for men (a) and women (b).

men and women using pooled samples from the 1994, 1995, 1996 March Current Population Surveys. The data represent mean log hourly earnings by single year of age for individuals with 10, 12 and 16 years of education. Plotted along with the actual means are the fitted values obtained from models like (1′) that include a cubic term in potential experience.[5] Comparisons of the fitted and actual data suggest that age-earnings profiles for US men and women are fairly smooth, and are reasonably well-approximated by a simple variant of the standard human capital earnings function. Nevertheless, even a cubic version of Mincer's model has some trouble fitting the precise curvature of the age profiles for different education groups in recent US data. In particular, the fitted models tend to understate the growth rate of earnings for younger college-educated men and women relative to high-school graduates, suggesting the need for more flexible interactions

[5] The samples include 102,718 men and 95,360 women age 16–66 with positive potential experience and average hourly earnings between \$2.00 and \$150.00 in 1995 dollars. Fifty-three percent of the sample have 10, 12, or 16 years of schooling and are used in graphs. The regression models are fit by gender to all education groups and include a linear education term, a cubic in experience, and a dummy variable for individuals of black race.

between education and experience. For some purposes these mis-specifications may not matter much. In other applications, however, biases in the fitted age profiles of different education groups may lead to serious misunderstandings.

2.2. *Measurement of education*

In addition to imposing separability between the effects of education and experience, the standard human capital earnings function dictates that log earnings are a *linear* function of years of completed education. There are two (related) hypotheses embedded in this specification: first, that the correct measure of education is the number of years of completed education; and second, that each additional year of schooling has the same proportional effect on earnings, holding constant years in the labor market. Assuming that these conditions are satisfied, the coefficient b in Eq. (1) completely summarizes the effect of education in the labor market. It is now conventional to refer to b as "the return to education".[6] As shown in Willis (1986, p. 532) if (1) or (1′) is correctly specified then b is in fact the internal rate of return to schooling investments, assuming that education is free and that students earn nothing while in school.

The use of years of completed education as a measure of schooling has a long history in the United States. Such data were collected in the 1940–1980 Decennial Censuses and in the Current Population Surveys from the 1940s to the early 1990s. Years of schooling has substantial face validity in the US education system, but is less natural in countries with multiple education streams (e.g., Germany or France) where high school graduation may entail different years of schooling depending on whether a student plans to go to university, vocational college, or start work right away.[7]

Even within the US many analysts have argued that credentials (such as a high school diploma or college degree) matter more than years of schooling per se. This hypothesis has come to be known as the "sheepskin effect" – the existence of wage premiums for fulfilling the final years of elementary school, high school, or college. Hungerford and Solon (1987) and Belman and Heywood (1991) augment a standard earnings function like (1) with variables to capture non-linearities at 8, 12, or 16 years of education. These authors find some evidence of non-linearity, especially around the 16th year of schooling (corresponding to college graduation).[8] Park (1994) analyzed a large sample of CPS data and concluded that most of the apparent non-linearity at 16 years of education arises from the relatively small difference in earnings between individuals with 14 and 15 years of schooling (i.e., an exceptionally low return to the 15th year of schooling, rather than an

[6] In fact, the education coefficient in any statistical model of wages (or earnings) is generally referred to as the "return to education", regardless of what other control variables are included in the model. This can lead to some confusion when age rather than potential experience (X) is included as a control, since the derivative of Eq. (1) with respect to schooling holding constant age is $b - c - 2dX$. Thus the "return to education" is generally lower in models that control for age rather than experience (Mincer, 1974, p. 84).

[7] Historically there were some inter-state differences in education systems in the US: for example, South Carolina had only three years of high school in the early 20th Century.

[8] See also Goodman (1979).

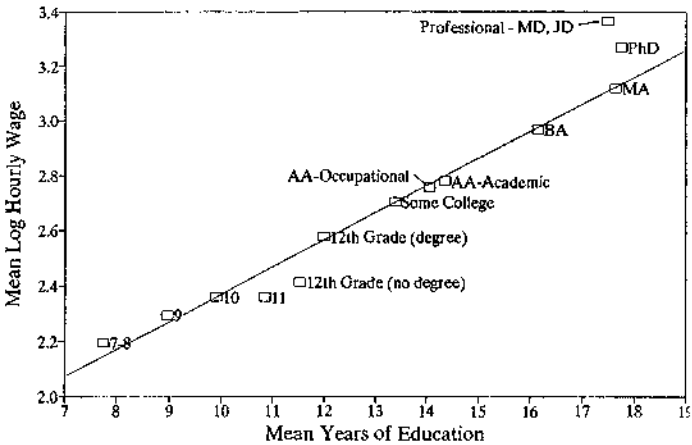

Fig. 2. Relationship between mean log hourly wages and completed education, men aged 40–45 in 1994–1996 Current Population Survey. Mean education by degree category estimated from February 1990 CPS.

exceptionally high return to the 16th year of schooling). Apart from this feature, Park shows that the linear functional form provides a surprisingly good fit to the data.

Despite economists' general satisfaction with the traditional measure of schooling, in the late 1980s the US Census Bureau decided to shift toward a degree-based system of measuring post-high-school education (see Kominski and Siegel, 1992). Thus, individuals in the 1990 Census and recent Current Population Surveys were no longer asked how many years of college they had completed: rather they were asked to report their college degrees. This change makes it more difficult to estimate the standard human capital earnings model with recent US data, or to measure changes in the structure of education-related wage differentials. Nevertheless, a concordance between the older years-of-education variable and the new degree-based variable can be constructed from a cross-tabulation of responses to the two questions included in a supplement to the February 1990 CPS. Use of this concordance provides some rather surprising support for the linearity assumption embedded in Mincer's original specification.[9]

Fig. 2 shows wage and schooling data for a sample of men age 40–55 in the 1994–1996 CPS.[10] Mean log wages for each education group (e.g., men with a junior college or Associates degree in an academic program, denoted by "AA-Academic" in the graph) are graphed against the mean number of years of education for the group measured in the February 1990 concordance. Apart from men who report 11 years of schooling, or 12 years

[9] See Park (1994, 1996) for further analysis of the linearity assumption.

[10] I use men in this age range to abstract from the effects of experience. As shown in Fig. 1a, after age 40 the age-earnings profiles of different education groups are roughly parallel.

with no high school degree, the data for individuals with between 7 and 18 years of education lie remarkably close to a line that joins the high school graduates and the college graduates (superimposed on the figure). The two highest-education groups are also off the line. My guess is that this reflects the censoring of the years-of-schooling variable, which was only reported to a maximum of 18 years.[11] Based on the patterns in Fig. 2, it may be reasonable to assign an estimate of the years of completed education to each reported education class and assume a linear functional form.

2.3. *Which measure of earnings?*

The literature on the human capital earnings function has analyzed a variety of earnings measures – annual, weekly, hourly – almost always in logarithmic form. The popularity of the log transformation reflects several factors. For one, the distribution of log earnings (especially log hourly wages) is surprisingly close to a normal distribution. Other things equal, many data analysts would therefore prefer to model the log of earnings. Another practical reason for using the log transformation is the apparent success of the standard (semi-logarithmic) human capital earnings function. As demonstrated in Fig. 1a,b, the distribution of log earnings across age and education groups is closely-approximated by the sum of a linear schooling term and a polynomial in experience. Conditional on the functional form of the right-hand side of Eq. (1), Heckman and Polachek (1974) investigated alternative transformations of earnings and concluded that the log transformation is the best in the Box–Cox class. Finally, and perhaps as important as any other consideration, the log transformation is convenient for interpretation.

The choice of time frame over which to measure earnings is often dictated by necessity: some datasets report annual earnings whereas others report hourly or weekly wages. Since individuals with higher schooling tend to work more, the measured return to schooling will be higher for weekly or annual earnings than for hourly earnings. This fact is illustrated in Table 1, which reports the estimated education coefficients from models analogous to Eq. (1′) fit to earnings and hours data for men and women in the 1994–1996 March CPS. The CPS questionnaire inquires about earnings last year, total weeks worked in the previous year, and usual hours per week last year. By construction,

$$\text{Annual earnings} = \text{Hourly Earnings} \times \text{Hours/Week} \times \text{Weeks}.$$

When log annual earnings are regressed on education and other controls, the estimated education coefficient is therefore the sum of the education coefficients for parallel models fit to the log of hourly earnings, the log of hours per week, and the log of weeks per year. In the US labor market in the mid-1990s, about two-thirds of the measured return to education observed in annual earnings data is attributable to the effect of education on earnings

[11] Individuals with a medical or law degree, for example, have at least 20 years of schooling, and many have more.

Table 1
Estimated education coefficients from standard human capital earnings function fit to hourly wages, annual earnings, and various measures of hours for men and women in March 1994–1996 Current Population Survey[a]

	Dependent variable				
	Log hourly earnings	Log hours per week	Log weeks per year	Log annual hours	Log annual earnings
	(1)	(2)	(3)	(4)	(5)
A. Men					
Education	0.100	0.018	0.025	0.042	0.142
coefficient	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
R-squared	0.328	0.182	0.136	0.222	0.403
B. Women					
Education	0.109	0.022	0.034	0.056	0.165
coefficient	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
R-squared	0.247	0.071	0.074	0.105	0.247

[a] Notes: Table reports estimated coefficient of linear education term in model that also includes cubic in potential experience and an indicator for non-white race. Samples include men and women age 16–66 who report positive wage and salary earnings in the previous year. Hourly wage is constructed by dividing wage and salary earnings by the product of weeks worked and usual hours per week. Data for individuals whose wage is under $2.00 or over $150.00 (in 1995 dollars) are dropped. Sample sizes are: 102,639 men and 95,309 women.

per hour, with the remainder attributable to the effects on hours per week and week per year.

2.4. Summary

This brief overview suggests that the human capital earnings function is alive and well. A simple regression model with a linear schooling term and a low-order polynomial in potential experience explains 20–35% of the variation in observed earnings data, with predictable and precisely-estimated coefficients in almost all applications. Close examination reveals that the model is too parsimonious to fully characterize the joint distribution of earnings, age and schooling. Nevertheless, it provides a natural starting point for building more complex models of earnings determination, and for investigating the effects of other covariates such as race, gender, and firm characteristics. Moreover, the conventional model serves as a useful benchmark for theorizing about the effects of education in the labor market. From this point of view, the approximate linearity of earnings with respect to schooling and the separability of the effects of education and experience are useful simplifications that can aid in the formulation of tractable theoretical models.

3. Causal modelling of the return to education

3.1. Theoretical issues

Most of the conceptual issues underlying the interpretation of recent studies of the return to education can be illustrated in the framework of a simple static model that builds on Becker (1967). According to this model, each individual faces a market opportunity locus that gives the level of earnings associated with alternative schooling choices. A static model abstracts from the dynamic nature of the schooling and earnings processes and focusses instead on the relationship between completed schooling and average earnings over the lifecycle. Such a focus is justified if people finish their formal schooling before entering the labor market (other than on a casual or part-time basis) and if the effect of schooling on log earnings is separable from the effect of experience, as is assumed in the standard human capital earnings function. In fact the transition from school to work is often a bumpy one, as young adults move back and forth between full-time or part-time enrollment and part-time or full-time work.[12] Nevertheless, most people have completed their formal schooling by their mid-20s.[13]

An analytically tractable version of Becker's model is developed in Card (1995a). Following that presentation, let $y(S)$ denote the average level of earnings (per year) an individual will receive if he or she acquires schooling level S.[14] Assume that an individual chooses S to maximize a utility function $U(S,y)$, where

$$U(S, y) = \log y - h(s), \tag{2}$$

and h is some increasing convex function. This function generalizes the discounted present value (DPV) objective function

$$\int_S^\infty y(S)\exp(rt)\mathrm{d}t = y(S)\exp(-rS)/r,$$

which is appropriate if individuals discount future earnings at a rate r, schooling is measured in years, and it is assumed that individuals earn nothing while in school and $y(S)$ per year thereafter. The DPV objective function sets $h(S) = rS$. More generally, however, $h(S)$ may be strictly convex if the marginal cost of each additional year of

[12] Angrist and Newey (1991) study the earnings changes associated with education increments acquired after young men enter the labor market on a full time basis.

[13] By age 24, fewer than one-fifth of US adults were enrolled in school (even on a part-time basis) in the early 1990s. A simple tabulation of enrollment rates by age suggests that the transition between school and work has become sharper over the past two decades, in the US at least. For example although enrollment rates of 20 year olds are now higher than in the late 1970s (47% enrolled in 1992 versus 37% in 1977) the enrollment rates of people in their late 20s are lower today (e.g., 7% for 30 year olds in 1992 versus 10% in 1977). These tabulations are from the October Current Population Survey and combine men and women.

[14] The market opportunity locus $y(S)$ may reflect productivity effects of higher education, and/or other forces such as signalling.

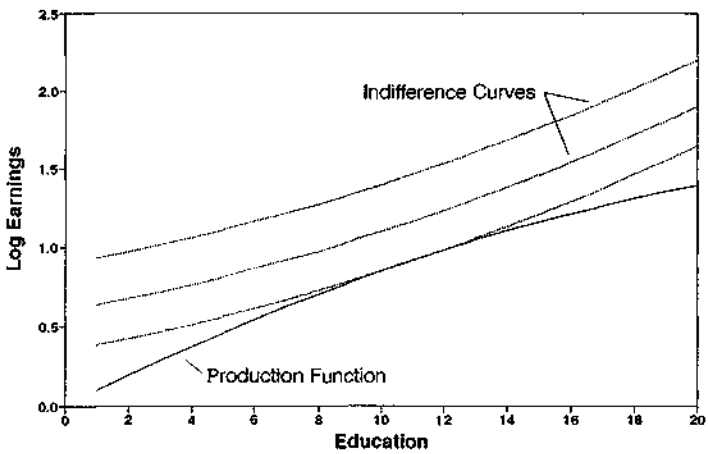

Fig. 3. Determination of optimum schooling.

schooling rises by more than the foregone earnings for that year, either because of credit market considerations (Becker, 1967) or taste factors.[15]

An optimal schooling choice satisfies the first-order condition

$$h'(S) = y'(S)/y(S),$$

as illustrated in Fig. 3. An important feature of the class of preference functions defined by Eq. (2) is linearity in log earnings. This means that the indifference curves in Fig. 3 are vertically parallel, with the immediate implication that any factor that raises log earnings for *all* levels of schooling has no effect on the optimal schooling choice. In principle this need not be true. For example, Griliches (1977) presents a variant of DPV preferences with the feature that a uniform upward shift in log earnings for all levels of schooling leads to a lower schooling choice.

Individual heterogeneity in the optimal schooling choice illustrated in Fig. 3 arises from two sources: differences in the costs of (or tastes for) schooling, represented by heterogeneity in $h(S)$; and differences in the economic benefits of schooling, represented by heterogeneity in the marginal return to schooling $y'(S)/y(S)$. A simple specification of these heterogeneity components is

$$y'(S)/y(S) = b_i - k_1 S, \tag{3a}$$

$$h'(S) = r_i + k_2 S, \tag{3b}$$

[15] Note that the marginal rate of substitution (MRS) between income and schooling is $y(S)h'(S)$. Under a DPV criterion MRS $= ry(S)$, since the opportunity costs of the Sth year schooling are just the foregone earnings $y(S)$. If $h'(S)$ is increasing in S, the MRS rises faster than $y(S)$.

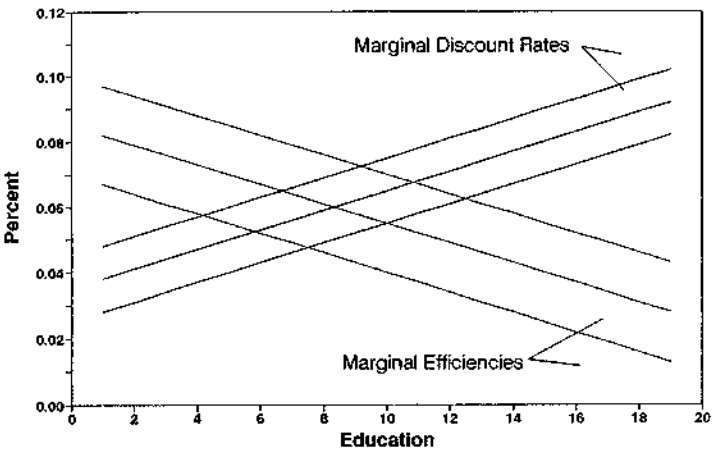

Fig. 4. Marginal benefit and marginal cost schedules for different individuals.

where b_i and r_i are random variables with means $\bar{b}$ and $\bar{r}$ and some joint distribution across the population $i = 1, 2, \ldots$, and k_1 and k_2 are non-negative constants. This specification implies that the optimal schooling choice is *linear* in the individual-specific heterogeneity terms,

$$S_i^* = (b_i - r_i)/k, \tag{4}$$

where $k = k_1 + k_2$. Fig. 4 illustrates the determination of optimal schooling using the marginal benefit and marginal cost schedules described by Eqs. (3a) and (3b).

Since formal schooling is usually completed early in life, individuals do not necessarily know the parameters of their earnings functions when they make their schooling choices. Thus, b_i should be interpreted as the individual's best estimate of his or her earnings gain per year of education, as of early adulthood. One might expect this estimate to vary less across individuals than their realized values of schooling. Moreover, the distribution of b_i may change over time with shifts in labor market conditions, technology, etc.[16] For simplicity, however, I will treat b_i as known at the beginning of the lifecycle and fixed over time: this assumption probably leads to some overstatement of the role of heterogeneity of b_i in the determination of schooling and earnings outcomes.

At the optimal level of schooling described by Eq. (4) individual i's marginal return to schooling is

$$\beta_i = b_i - k_1 S_i^* = b_i(1 - k_1/k) + r_i k_1/k.$$

Even in this very simple model equilibrium entails a *distribution* of marginal returns

[16] If changes over time cause the mean return $\bar{b}$ for a cohort to rise or fall, but leave the distribution of b_i otherwise unaffected, then the results presented below are unaffected.

across the population unless one of two conditions is satisfied: (a) $r_i = \bar{r}$ for all i and $k_2 = 0$ (i.e., linear indifference curves with a uniform slope $\bar{r}$ in Fig. 3); or (b) $b_i = \bar{b}$ for all i and $k_1 = 0$ (i.e., linear opportunity locuses with a uniform slope $\bar{b}$ in Fig. 3).

In general equilibrium the distribution of marginal returns to schooling is endogenous: a greater supply of highly-educated workers will presumably lower $\bar{b}$, and might also affect other characteristics of the distribution of b_i.[17] From the point of view of a cohort of young adults deciding on their education, however, the distribution of returns to education is arguably exogenous. I therefore prefer to interpret Eq. (4) as a partial equilibrium description of the relative education choices of a cohort of young adults, given their family backgrounds and the institutional environment and economic conditions that prevailed during their late teens and early 20s. Differences across cohorts in these background factors will lead to further variation in the distribution of marginal returns to education in the population as a whole.

3.2. Observed schooling and earnings outcomes

To understand the implications of the preceding model for observed schooling and earnings outcomes, note that Eq. (3a) implies a model for log earnings of the form

$$\log y_i = \alpha_i + b_i S_i - \frac{1}{2} k_1 S_i^2,$$

where α_i is a person-specific constant of integration. This is a somewhat more general version of the semi-logarithmic functional form adopted in Mincer (1974) and hundreds of subsequent studies. In particular, individual heterogeneity potentially affects both the *intercept* of the earnings equation (via α_i) and the *slope* of the earnings-schooling relation (via b_i). It is convenient to rewrite this equation as

$$\log y_i = a_0 + \bar{b} S_i - \frac{1}{2} k_1 S_i^2 + a_i + (b_i - \bar{b}) S_i, \tag{5}$$

where $a_i \equiv \alpha_i - a_0$ has mean 0. Eqs. (4) and (5) together describe a two-equation system for schooling and earnings in terms of the underlying random variables a_i, b_i, and r_i.

To proceed, consider the linear projections of a_i and $(b_i - \bar{b})$ on observed schooling:

$$a_i = \lambda_0 (S_i - \bar{S}) + u_i, \tag{6a}$$

$$b_i - \bar{b} = \psi_0 (S_i - \bar{S}) + v_i, \tag{6b}$$

where $\bar{S}$ represents the mean of schooling and $\mathrm{E}[S_i u_i] = \mathrm{E}[S_i v_i] = 0$. The parameters λ_0 and ψ_0 in Eqs. (6a) and (6b) are theoretical regression coefficients:

[17] See Freeman (1986) and Willis (1986) for some discussion of the general equilibrium implications of optimal schooling models.

$$\lambda_0 = \frac{\text{cov}(a_i, S_i)}{\text{var}(S_i)} = k\frac{\sigma_{ba} - \sigma_{ra}}{\sigma_b^2 + \sigma_r^2 - 2\sigma_{br}}$$

and

$$\psi_0 = \frac{\text{cov}(b_i, S_i)}{\text{var}(S_i)} = k\frac{\sigma_b^2 - \sigma_{br}}{\sigma_b^2 + \sigma_r^2 - 2\sigma_{br}},$$

where σ_b^2, σ_r^2, and σ_{br} denote the variances and covariance of b_i and r_i, and σ_{ba} and σ_{ra} denote the covariances of b_i and r_i with a_i. For simplicity, assume that b_i and r_i have a jointly symmetric distribution.[18] Then, using Eq. (A.3), and the fact that a linear projection of S_i^2 on S_i has slope $2\bar{S}$, it is readily shown that the probability limit of the ordinary least squares (OLS) regression coefficient b_{ols} from a regression of log earnings on schooling is

$$\text{plim} b_{\text{ols}} = \bar{b} + \lambda_0 - k_1\bar{S} + \psi_0\bar{S} = \bar{\beta} + \lambda_0 + \psi_0\bar{S}, \tag{7}$$

where $\bar{\beta} \equiv \text{E}[\beta_i] = \text{E}[b_i - k_1 S_i] = \bar{b} - k_1\bar{S}$ is the average marginal return to schooling in the population.[19]

Eq. (7) generalizes the conventional analysis of ability bias in the relationship between schooling and earnings (see Griliches, 1977).[20] Suppose that there is no heterogeneity in the marginal benefits of schooling (i.e., $b_i = \bar{b}$) and that log earnings are linear in schooling (i.e., $k_1 = 0$). In this case (7) implies that

$$\text{plim} b_{\text{ols}} - \bar{b} = \lambda_0.$$

This is the standard expression for the asymptotic bias in the estimated return to schooling that arises by applying the "omitted variables" formula to an earnings model with a constant schooling coefficient $\bar{b}$. According to the model presented here, this bias arises through the correlation between unobserved ability a_i and the marginal cost of schooling r_i.[21] If marginal costs are lower for children from more privileged family backgrounds, and if these children would also tend to earn more at any level of schooling, then $\sigma_{ra} < 0$, implying that $\lambda_0 > 0$.

If both the intercept and slope of the earnings function vary across individuals then the situation is more complicated. Since people with a higher return to education will tend to acquire more schooling, a cross-sectional regression of earnings on schooling yields an upward-biased estimate of the average marginal return to schooling, even ignoring variation in the intercepts of the earnings function. The magnitude of this endogeneity or self-

[18] This assumption implies that $\text{E}[(b_i - \bar{b})^3] = \text{E}[(r_i - \bar{r})^3] = \text{E}[(r_i - \bar{r})(b_i - \bar{b})^2] = \cdots = 0$.

[19] If the random variables r_i and b_i are not symmetrically distributed then Eq. (7) contains an additional term equal to $\text{E}[(b_i - \bar{b})(S_i - \bar{S})^2]$. See Appendix A.

[20] Throughout this paper I use the term "bias" to refer to the difference between the probability limit of an estimator and some target parameter: typically the average marginal return to schooling in the population under study.

[21] As noted earlier, the form of Eq. (2) rules out a direct connection between a_i and optimal schooling choice.

selection bias $\psi_0\bar{S}$ depends on the importance of variation in b_i in determining the overall variance of schooling outcomes.

To see this, note that the variance of schooling is $(\sigma_b^2 + \sigma_r^2 - 2\sigma_{br})/k^2$. The fraction of the variance of schooling attributable to differences in the slope of the earnings-schooling relation (as opposed to differences in tastes or access to funds) can be defined as

$$f = \frac{\sigma_b^2 - \sigma_{br}}{\sigma_b^2 + \sigma_r^2 - 2\sigma_{br}}.$$

Assuming that $\sigma_{br} \leq 0$ (i.e., that the marginal benefits of schooling are *no higher* for people with higher marginal costs of schooling), this "fraction" is bounded between 0 and 1. The auxiliary regression coefficient defined in Eq. (6b) is $\psi_0 = kf \geq 0$. Thus, the endogeneity bias component in b_{ols} is

$$\psi_0\bar{S} = kf\bar{S} \geq 0.$$

Even ignoring the traditional ability bias term λ_0, b_{ols} is therefore an *upward-biased* estimator $\bar{\beta}$; moreover, the greater is f, the greater is the endogeneity bias.

Superficially, the earnings model specified by Eq. (5) seems inconsistent with the observation that the cross-sectional relationship between log earnings and schooling is approximately linear. Because of the endogeneity of schooling, however, S_i and $(b_i - \bar{b})$ are positively correlated across the population, leading to a *convex* relationship between log earnings and schooling in the absence of any concavity in the underlying opportunity locuses. More formally, substitution of (6a) and (6b) into Eq. (5) leads to

$$\begin{aligned}\log y_i &= a_0 + \bar{b}S_i - 1/2k_1S_i^2 + \lambda_0(S_i - \bar{S}) + \psi_0 S_i(S_i - \bar{S}) + u_i + S_i v_i \\ &= c + (\bar{b} + \lambda_0 - \psi_0\bar{S})S_i + (\psi_0 - 1/2k_1)S_i^2 + u_i + S_i v_i, \qquad (5')\end{aligned}$$

where c is a constant. If $\mathrm{E}[u_i|S_i] = \mathrm{E}[v_i|S_i] = 0$ (assumptions which are somewhat stronger than the orthogonality conditions implicit in Eqs. (6a) and (6b)), then Eq. (5′) implies that $\mathrm{E}[\log y_i|S_i]$ is a quadratic function of schooling with second-order coefficient $(\psi_0 - 1/2k_1)$. The empirical relationship between log earnings and schooling will therefore be approximately linear if and only if $k_1 \approx 2\psi_0$. The bigger is the contribution of variation in b_i to the overall variance of schooling, the larger is ψ_0 and the more convex is the observed relationship between log earnings and schooling.[22]

3.3. Measurement error

An important issue in the literature on returns to schooling is the effect of survey measurement error in schooling. As emphasized by Griliches (1977, 1979) measurement errors in

[22] The observation that the cross-sectional relationship between log earnings and schooling is approximately linear should not be pushed too far. Given the dispersion in residual earnings, a quadratic function of schooling with a non-trivial second order term may well appear linear over the limited range of school outcomes actually observed in any sample.

schooling would be expected to lead to a downward bias in any OLS estimator of the relationship between schooling and earnings. A conventional assumption is that *observed* schooling (S_i^o) differs from true schooling (S_i) by an additive error

$$S_i^o = S_i + \varepsilon_i,$$

with $E[\varepsilon_i] = 0$, $E[S_i\varepsilon_i] = 0$, and $E[\varepsilon_i^2] = \sigma_\varepsilon^2$. Assuming that Eq. (7) describes the probability limit of an OLS estimator using *true* schooling, the use of observed schooling will yield an OLS estimator with

$$\text{plim}(b_{\text{ols}}) = R_0\{\bar{\beta} + \lambda_0 + \psi_0\bar{S}\}, \tag{8}$$

where

$$R_0 \equiv \text{cov}[S_i^o, S_i]/\text{var}[S_i^o] = \text{var}[S_i]/\{\text{var}[S_i] + \sigma_\varepsilon^2\}$$

is the reliability of S_i^o, or the signal-to-total-variance ratio of observed schooling. Treating b_{ols} as an estimator of $\bar{\beta}$, the asymptotic bias is

$$\text{Bias}_{\text{ols}} = R_0(\lambda_0 + \psi_0\bar{S}) - (1 - R_0)\bar{\beta}.$$

Research over the past three decades has generally found that the reliability of self-reported schooling is about 90%,[23] suggesting that the second term in this expression is on the order of $-0.1\bar{\beta}$ in most datasets. Depending on the magnitudes of λ_0 and $\psi_0\bar{S}$, this may partially offset the presumably positive biases imparted by the correlations between schooling and the ability components a_i and b_i.

The preceding argument hinges on the assumption that measurement errors in schooling are uncorrelated with true schooling. Since schooling is typically measured as a discrete variable with outcomes ranging between fixed upper and lower limits, however, the errors in reported schooling are probably mean-regressive.[24] Specifically, individuals with very high levels of schooling *cannot* report positive errors in schooling, whereas individuals with very low levels of schooling *cannot* report negative errors in schooling. If the errors in observed schooling measures are negatively correlated with true schooling, the actual reliability of an observed schooling measure may be slightly higher than the estimated reliability inferred from the correlation between two alternative measures of schooling.[25]

[23] See, e.g., Siegel and Hodge (1968), Miller et al. (1995), and Ashenfelter and Rouse (1998). Interestingly, the very limited available evidence on administrative measures of schooling suggests a similar reliability ratio; e.g., Kane et al. (1997); Isacsson (1997).

[24] This point is raised in a recent paper by Kane et al. (1997).

[25] To see this, suppose that there are two measures x_1 and x_2 of a true quantity x, with $x_j = x + e_j$, and assume that $E[e_j|x] = -\alpha(x - \mu)$, for $j = 1, 2$, where μ is the mean of x. Decompose the measurement errors as $e_j = -\alpha(x - \mu) + v_j$, and assume that the v_j's are independent of each other and x, and have equal variances. The reliability of x_1 is $R = \text{cov}[x, x_1]/\text{var}[x_1]$. Traditionally, reliability is measured by $\rho = \text{cov}[x_1, x_2]/\text{var}[x_1]$ (assuming that x_1 and x_2 have the same variance). It is straightforward to show that $\rho = (1 - \alpha)R$.

3.4. Instrumental variables estimates of the return to schooling

Social scientists have long recognized that the cross-sectional correlation between education and earnings may differ from the true causal effect of education. A standard solution to the problem of causal inference is instrumental variables (IV): a researcher posits the existence of an observable covariate that affects schooling choices but is uncorrelated with (or independent of) the ability factors a_i and b_i. For example, suppose that the marginal cost component r_i is linearly related to a set of variables Z_i:

$$r_i = Z_i\pi_1 + \eta_i.$$

In this case the school choice equation becomes

$$S_i = Z_i\pi + (b_i - \eta_i)/k = \pi_0 + Z_i\pi + \xi_i, \tag{4'}$$

where $\pi = -\pi_1/k$ and $\xi_i \equiv (b_i - \bar{b} - \eta_i)/k$. In the recent literature much attention has focussed on what might be called institutional sources of variation in schooling, attributable to such features as the minimum school leaving age, tuition costs for higher education, or the geographic proximity of schools. Such institutional factors stand a reasonable chance of satisfying the strict exogeneity assumptions required for a legitimate instrumental variable.

In the presence of heterogeneous returns to education the conditions required to yield an interpretable IV estimator are substantially stronger than those required when the only source of ability bias is random variation in the constant of the earnings equation (i.e., variation in a_i).[26] Wooldridge (1997) presents a useful analysis that can be directly applied to the system of Eqs. (4') and (5). Assume for the moment that $k_1 = 0$ in the earnings equation, and consider three additional assumptions on the unobservable components of (4') and (5):

$$\mathrm{E}[\eta_i \mid Z_i] = 0, \qquad \mathrm{E}[a_i \mid Z_i] = 0, \qquad \mathrm{E}[(b_i - \bar{b}) \mid Z_i] = 0, \tag{9a}$$

$$\mathrm{E}[(b_i - \bar{b})^2 \mid Z_i] = \sigma_b^2, \tag{9b}$$

$$\mathrm{E}[\xi_i \mid b_i, Z_i] = \rho_1(b_i - \bar{b}). \tag{9c}$$

Eq. (9a) specifies that the individual-specific heterogeneity components are all mean independent of the instrument Z. Eq. (9b) states that the second moment of b_i is also conditionally independent of Z_i. Finally, Eq. (9c) states that the conditional expectation of the unobserved component of optimal school choice (ξ_i) is linear in b_i. Since $\xi_i \equiv (b_i - \bar{b} - \eta_i)/k$, a sufficient condition for (9c) is that $\mathrm{E}[\eta_i|b_i, Z_i] = \rho(b_i - \bar{b})$, in

[26] If the only individual-specific component of ability is a_i then Eqs. (4') and (5) constitute a standard simultaneous equations system and one need only assume $\mathrm{E}[a_iZ_i] = \mathrm{E}[\eta_iZ_i] = 0$. The interpretation of IV in the presence of random coefficients is pursued in a series of papers by Angrist and Imbens (1995) and Angrist et al. (1996). Heckman and Vytlacil (1998) present some results similar to those discussed here.

which case $\rho_1 = (1 - \rho)/k$. This will be true if b_i and η_i have a bivariate normal distribution that is independent of Z_i, for example.

Under assumptions (9a)–(9c), the conditional expectation of the residual earnings component attributable to heterogeneity in b_i is

$$\mathrm{E}[(b_i - \bar{b})S_i \mid Z_i] = \mathrm{E}[\mathrm{E}[(b_i - \bar{b})S_i \mid b_i, Z_i] \mid Z_i] = \mathrm{E}[(b_i - \bar{b})\mathrm{E}[S_i \mid b_i, Z_i] \mid Z_i]$$

$$= \mathrm{E}[(b_i - \bar{b})\mathrm{E}[Z_i\pi + \xi_i \mid b_i, Z_i] \mid Z_i] = \rho_1\sigma_b^2.$$

It follows that

$$\mathrm{E}[\log y_i \mid Z_i] = a_0 + \bar{b}Z_i\pi + \rho_1\sigma_b^2.$$

Thus, the use of Z_i as an instrument for education will lead to a consistent estimate of the mean return to schooling $\bar{b}$ (but an inconsistent estimate of a_0).[27] If earnings are a quadratic function of schooling (i.e., $k_1 > 0$) Wooldridge notes that the squared predicted value of schooling from Eq. (4′) can be added to the list of conditioning variables and the previous argument remains valid.

A closely-related alternative to IV estimation of a random coefficients model is a control function approach, first proposed in the schooling context by Garen (1984). In place of Eqs. (9b) and (9c), assume that the conditional expectations of a_i and b_i are linear in S_i and Z_i:

$$\mathrm{E}[a_i \mid S_i, Z_i] = \lambda_1 S_i + \lambda_2 Z_i, \tag{10a}$$

$$\mathrm{E}[b_i - \bar{b} \mid S_i, Z_i] = \psi_1 S_i + \psi_2 Z_i, \tag{10b}$$

As noted in Appendix A, maintaining the assumptions that $\mathrm{E}[a_i|Z_i] = \mathrm{E}[b_i - \bar{b}|Z_i] = 0$, these conditions are equivalent to assuming

$$\mathrm{E}[a_i \mid S_i, Z_i] = \lambda_1\xi_i, \tag{10a′}$$

$$\mathrm{E}[b_i - \bar{b} \mid S_i, Z_i] = \psi_1\xi_i, \tag{10b′}$$

where ξ_i is defined in Eq. (4′). It follows immediately that

$$\mathrm{E}[\log y_i \mid S_i, Z_i] = a_0 + \bar{b}S_i - 1/2k_1S_i^2 + \lambda_1\xi_i + \psi_1\xi_iS_i. \tag{11}$$

The control function approach to estimation of the average return to schooling is to substitute the estimated residual $\hat{\xi}_i$ from the reduced form schooling Eq. (4′) in place of ξ_i in Eq. (11). Note that the inclusion of $\hat{\xi}_i$ as an additional regressor in the earnings function is numerically equivalent to IV using Z_i as an instrument for S_i. Under the assumption that $\mathrm{E}[a_i|Z_i] = 0$ the addition of $\hat{\xi}_i$ to the estimated earnings function purges

[27] Assumptions (9a) and (9b) are not the only ones that lead to a consistent IV estimator. Wooldridge proposes as an alternative the pair of assumptions: $\mathrm{E}[\xi_i^2|Z_i] = \sigma_\xi^2$ and $\mathrm{E}[(b_i - \bar{b})|\xi_i, Z_i] = \tau\xi_i$. The proof of consistency of the IV estimator then proceeds by noting that $\mathrm{E}[(b_i - \bar{b})S_i|Z_i] = \mathrm{E}[\mathrm{E}[(b_i - \bar{b})S_i|\xi_i, Z_i]|Z_i] = \tau\sigma_\xi^2$.

the effect of a_i on the observed relationship between log earnings and schooling. In general, however, standard IV will *not* eliminate the influence of b_i on the covariance between schooling and earnings, unless $E[(b_i - \bar{b})S_i|Z_i]$ is independent of Z_i (as is the case under Wooldridge's assumptions). Under assumption (10b) (or equivalently (10b')), the addition of $\check{\xi}_i S_i$ as a second control variable is sufficient to eliminate the endogeneity bias arising from the correlation between b_i and S_i. Thus, the control function approach might be viewed as a generalization of instrumental variables.

3.5. Limitations of instrumental variables

In the absence of assumptions such as those underlying Eqs. (9) or (10), even an instrumental variables estimator based on an exogenous instrument will not necessarily yield an asymptotically unbiased estimate of the average return to education. To illustrate this point, consider IV estimation using the change in education associated with a "schooling reform" that leads to a proportional reduction in the marginal cost of schooling for students in a specific set of schools (or in a specific cohort). Assume that the joint distribution of abilities and tastes (a_i, b_i, r_i) is the same for individuals who attended the reformed schools (indexed by $Z_i = 1$) and those who did not (indexed by $Z_i = 0$), but that in the reformed schools the optimal school choice is given by

$$S_i^* = (b_i - \theta r_i)/k, \tag{4''}$$

where $0 < \theta < 1$. Clearly, differences in Z_i will be associated with differences in average levels of schooling. Moreover, by assumption the distributions of ability are the same among students who attended the two sets of schools. In this setting, however, the treatment effect of the school reform is larger for individuals who would have had lower schooling levels in the absence of the reform, causing potential difficulties for the interpretation of an IV estimator based on Z_i.

Let $r_i = \bar{r} + \eta_i$, and observe that among the comparison group of individuals who attend the unreformed schools,

$$S_i = (\bar{b} - \bar{r})/k + (b_i - \bar{b} - \eta_i)/k = \pi_0 + \xi_{i0},$$

whereas among the treatment group of individuals who attended reformed schools,

$$S_i = (\bar{b} - \theta\bar{r})/k + (b_i - \bar{b} - \theta\eta_i)/k = \pi_1 + \xi_{i1}.$$

Assume that $E[\eta_i|b_i] = \rho(b_i - \bar{b})$. Then

$$E[\xi_0 \mid b_i] = \rho_0(b_i - \bar{b}),$$

where $\rho_0 = (1 - \rho)/k$, whereas

$$E[\xi_1 \mid b_i] = \rho_1(b_i - \bar{b}),$$

where $\rho_1 = (1 - \theta\rho)/k$. Thus, the correlation between the reduced form schooling error and unobserved ability is *different* in the treatment and control groups, leading to a viola-

tion of the assumptions needed for IV or a control function estimator to yield a consistent estimate of the average marginal return to schooling.

The school reform causes a given individual (characterized by the triplet (a_i,b_i,η_i)) to increment his or her schooling by an amount

$$\Delta S_i = \pi_1 - \pi_0 + \eta_i(1-\theta)/k.$$

The (first-order) effect on this individual's earnings is

$$\Delta \log y_i = \beta_i \Delta S_i,$$

where β_i is i's marginal return to schooling in the absence of the intervention:

$$\beta_i = \bar{\beta} + b_i - \bar{b} - k_1(S_i - \bar{S}) = \bar{\beta} + (b_i - \bar{b})(1 - k_1/k) + \eta_i k_1/k.$$

Using these expressions, the expected earnings differential between individuals in the treatment group and the control group is

$$\mathrm{E}[\Delta \log y_i] = \bar{\beta}(\pi_1 - \pi_0) + k_1/k^2(1-\theta)\sigma_\eta^2 + \sigma_{b\eta}(1-\theta)(1-k_1/k)/k,$$

where expectations are taken with respect to the joint distribution of (a_i,b_i,η_i). The IV estimator of the return to schooling based on the instrument Z_i, b_{iv}, has probability limit

$$\begin{aligned}\mathrm{plim} b_{\mathrm{iv}} &= \frac{\mathrm{E}[\log y_i \mid Z_i = 1] - \mathrm{E}[\log y_i \mid Z_i = 0]}{\mathrm{E}[S_i \mid Z_i = 1] - \mathrm{E}[S_i \mid Z_i = 0]} \\ &= \bar{\beta} + \frac{1-\theta}{k(\pi_1 - \pi_0)}\{\sigma_\eta^2 k_1/k + \sigma_{b\eta}(1 - k_1/k)\}.\end{aligned}$$

Note that if η_i is constant for all i (in which case everyone gets the same increment to schooling), then $\sigma_\eta^2 = \sigma_{b\eta} = 0$, and the IV estimator is consistent for $\bar{\beta}$. Otherwise, assuming that $\sigma_{b\eta} \leq 0$, so that individuals with higher returns to schooling have higher tastes for schooling or lower discount rates, the IV estimator may be positively or negatively biased relative to $\bar{\beta}$. A positive bias arises because the marginal return to schooling is decreasing in education if $k_1 > 0$: thus people with initially higher marginal costs of schooling tend to have higher marginal returns to an additional year of schooling. Lang (1993) labelled this phenomenon "discount rate bias". On the other hand, a negative bias arises because people with higher marginal costs of education, who are most affected by the school reform, have lower marginal returns to schooling if $\sigma_{b\eta} < 0$. The positive bias is more likely to dominate, the smaller is $|\sigma_{b\eta}|$ relative to σ_η^2 and the more concave are individual earnings functions.

To generalize this analysis slightly, suppose that the population can be divided into discrete subgroups of individuals ($g = 1, 2, \ldots$) who share common values for the latent ability and cost terms (a_g,b_g,η_g). Consider an intervention (such as a change in the compulsory schooling age) that leads to a change ΔS_g in the mean schooling of group g, and let β_g denote the marginal return to schooling for group g in the absence of the intervention. Finally, suppose that the intervention affects a treatment group of students

who are otherwise identical to those in a comparison group. In particular, assume that individuals in the treatment group and comparison group with the same latent ability and cost terms would have the same education and earnings in the absence of the intervention, and that the joint distributions of abilities and costs are the same in the two groups. Then an IV estimator of the return to schooling based on an indicator for treatment group status will have probability limit

$$\text{plim} b_{\text{iv}} = \frac{\text{E}[\beta_g \Delta S_g]}{\text{E}[\Delta S_g]},$$

where expectations are taken with respect to the probability distribution of the population across cells.[28] Note that if $\Delta S_g \geq 0$ for all g (which need not be true) then this expression can be interpreted as a weighted average of the marginal returns to education for each group, with weight ΔS_g.[29] A necessary and sufficient condition for $\text{plim} b_{\text{iv}} = \bar{\beta}$ is $\text{E}[\beta_g \Delta S_g] = \text{E}[\beta_g]\text{E}[\Delta S_g]$. Among the sufficient conditions for this equality are: (a) $\beta_g = \bar{\beta}$ (identical marginal returns for all groups); or (b) $\text{E}[\Delta S_g | \beta_g] = \Delta S$ (a homogeneous additive treatment effect of the schooling reform). In general, however, if there is some heterogeneity is the distribution of marginal returns to schooling, IV based on an intervention that affects a narrow subgroup of the population may lead to an estimated return to schooling above or below an OLS estimator for the same sample.

Two other aspects of the instrumental variables estimator are worth emphasizing. First, the probability limit of the IV estimator is unaffected by measurement error in schooling.[30] This in itself will lead to tendency for an IV estimator to exceed the corresponding OLS estimator of the effect of schooling on earnings. Second, the validity of a particular IV estimator depends crucially on the assumption that the instruments are uncorrelated with other latent characteristics of individuals that may affect their earnings. In the case of an IV estimator based on an indicator variable Z_i, for example, the IV estimator is numerically equal to the difference in mean log earnings between the $Z_i = 1$ group and the $Z_i = 0$ group, divided by the corresponding difference in mean schooling.[31] If the difference in schooling is small, *even minor differences* in mean earnings between the two groups will be blown up by the IV procedure. If Z_i were randomly assigned, as in a true experiment, this would not be a particular problem. In the case of quasi or natural experiments, however, inferences are based on difference between groups of individuals who attended schools at different times, or in different locations, or had differences in other characteristics such as month of

[28] This analysis can be generalized by allowing the latent variables to have different distributions among the treatment and comparison groups. This can be handled in principle by "reweighting" the comparison group, although the weights may not be directly observable.

[29] If ΔS_g is dichotomous (so that the change in schooling is either zero or a one unit effect) then the preceding analysis can be placed in the "local average treatment effect" framework developed by Angrist and Imbens (1995). See also Angrist et al. (1996).

[30] This assumes that the instrumental variable is uncorrelated with the measurement error in schooling.

[31] If other covariates are included in the model then the means for each subsample are adjusted for the effects of the covariates.

birth. The use of these differences to draw causal inferences about the effect of schooling requires careful consideration of the maintained assumption that the groups are otherwise identical.

3.6. *Family background*

While some of the most innovative recent research on the value of schooling has used institutional features of the education system to identify the causal effect of schooling, there is a long tradition of using family background information – such as mother's and father's education – to either directly control for unobserved ability or as an instrumental variable for completed education.[32] Interest in family background is driven by the fact that children's schooling outcomes are very highly correlated with the characteristics of their parents, and in particular with parents' education.[33] The strength of this correlation is illustrated in Table 2, which reports estimated coefficients from a simple regression of completed education on father's and mother's education, using samples of adult household heads from the 1972–1996 General Social Survey (GSS).[34] For a variety of subsamples, each additional year of schooling of either parent raises completed education by about 0.2 years, while a rise of 1 year in the parent's average education raises completed schooling by about 0.4 years. Roughly 30% of the observed variation in education among US adults is explained by parental education.[35]

Despite the strong intergenerational correlation in education, it is far from clear that family background measures are legitimate instrumental variables for completed education, even if family background has no independent causal effect on earnings. To illustrate this point, assume for the moment that there is no heterogeneity in the return to education (i.e., $b_i = \bar{b}$) and ignore any concavity in the log earnings function (i.e., assume $k_1 = 0$). In this case Eq. (5) becomes

$$\log y_i = a_0 + \bar{b}S_i + a_i, \tag{5''}$$

Consider a linear projection of the unobserved ability component on individual schooling and some measure of family background (F_i):

$$a_i = \lambda_1(S_i - \bar{S}) + \lambda_2(F_i - \bar{F}) + u_i', \tag{12}$$

This bivariate projection can be compared to the projection of a_i on S_i alone (i.e., Eq. (6a)) by considering two other auxiliary regressions

[32] Griliches (1979) presents a survey of research on family-based models of education and earnings.

[33] See Siebert (1985) for references to some of the literature on family background and education. Ashenfelter and Rouse (1998) show that up to 60% of the cross-sectional variation in schooling outcomes in their twins sample can be explained by (observable and unobservable) family factors.

[34] The models reported in Table 2 include controls for the age and birth year of the respondents, although the estimated coefficients (and R-squared coefficients) are not much different without these controls.

[35] The results in Table 2 are fairly typical of those found in the literature using other samples. If family background is measured by only one parent's education, the coefficient is generally in the range of 0.3–0.4.

Table 2
Effects of parental education on completed schooling[a]

		Father's education	Mother's education	*R*-squared
By race and gender				
1.	White men	0.23	0.20	0.26
	(N = 7330)	(0.01)	(0.01)	
2.	White women	0.20	0.21	0.32
	(N = 8547)	(0.01)	(0.01)	
3.	Black men	0.18	0.22	0.33
	(N = 705)	(0.03)	(0.04)	
4.	Black women	0.09	0.22	0.28
	(N = 1030)	(0.02)	(0.03)	
Men (all races) by birth cohort				
5.	Born before 1920	0.25	0.22	0.23
	(N = 430)	(0.05)	(0.05)	
6.	Born 1920–1934	0.26	0.24	0.22
	(N = 1590)	(0.03)	(0.03)	
7.	Born 1935–1944	0.24	0.24	0.26
	(N = 1785)	(0.02)	(0.02)	
8.	Born 1945–1954	0.22	0.19	0.23
	(N = 2482)	(0.02)	(0.02)	
9.	Born 1955–1964	0.26	0.11	0.23
	(N = 1593)	(0.02)	(0.02)	
Women (all races) by birth cohort				
10.	Born before 1920	0.21	0.25	0.29
	(N = 492)	(0.04)	(0.04)	
11.	Born 1920–1934	0.19	0.25	0.28
	(N = 1936)	(0.02)	(0.02)	
12.	Born 1935–1944	0.17	0.23	0.25
	(N = 2112)	(0.02)	(0.02)	
13.	Born 1945–1954	0.19	0.18	0.25
	(N = 2911)	(0.01)	(0.02)	
14.	Born 1955–1964	0.20	0.20	0.26
	(N = 1960)	(0.01)	(0.02)	

[a] Notes: Dependent variable in all models is years of completed education. Samples include individuals age 24–64 in the 1972–1996 General Social Survey with valid information on their own and both parents' education. Models in rows 1–4 include quadratic functions of respondent's age and birth year, in addition to father's and mother's education. Models in rows 5–14 include only a linear age term.

$$F_i = \delta_0 + \delta_s S_i + e_{1i}, \tag{13a}$$

$$S_i = \pi_0 + \pi_F F_i + e_{2i}, \tag{13b}$$

where e_{1i} is orthogonal to S_i and e_{2i} is orthogonal to F_i. The conventional omitted variables

formula implies that the coefficients in Eqs. (6a) and (12) are related by

$$\lambda_0 = \lambda_1 + \lambda_2\delta_s.$$

Moreover, δ_s and π_F are related to the correlation coefficient between S_i and F_i (ρ_{SF}) by

$$\delta_s\pi_F = \rho_{SF}^2.$$

Using these results it is possible to compare three potential estimators of Eq. (5″): the OLS estimator from a univariate regression of earnings on schooling (b_{ols}); the OLS estimator from a bivariate regression of earnings on schooling and family background (b_{biv}); and the IV estimator using F_i as an instrument for S_i (b_{iv}). The probability limits of these three estimators are

$$\text{plim}b_{\text{ols}} = \bar{b} + \lambda_0 = \bar{b} + \lambda_1 + \lambda_2\rho_{SF}^2/\pi_F,$$

$$\text{plim}b_{\text{biv}} = \bar{b} + \lambda_1,$$

$$\text{plim}b_{\text{iv}} = \text{cov}[\log y_i, F_i]/\text{cov}[S_i, F_i] = \bar{b} + \lambda_1 + \lambda_2/\pi_F.$$

In addition, the probability limit of the coefficient on F_i in the bivariate regression is just λ_2.[36] Assuming that $\lambda_1 \geq 0$, $\lambda_2 \geq 0$, and $\pi_F > 0$,

$$\bar{b} \leq \text{plim}b_{\text{biv}} \leq \text{plim}b_{\text{ols}} \leq \text{plim}b_{\text{iv}}.$$

If a_i and S_i are uncorrelated, controlling for F_i, then $\lambda_1 = 0$ and the bivariate OLS estimator is consistent for $\bar{b}$. Otherwise all three estimators are likely to be upward biased, with bigger biases in the univariate OLS and IV estimators than in the bivariate estimator unless $\lambda_2 = 0$.[37]

This analysis is readily extended to the case in which b_i varies across individuals. Assume that earnings are given by Eq. (5) and consider the projection of b_i on S_i and F_i:

$$b_i - \bar{b} = \psi_1(S_i - \bar{S}) + \psi_2(F_i - \bar{F}) + v_i'. \tag{14}$$

As with the coefficients λ_0 and λ_1, the coefficients ψ_0 in Eq. (6b) and ψ_1 in Eq. (14) are related by

$$\psi_0 = \psi_1 + \psi_2\delta_s = \psi_1 + \psi_2\rho_{SF}^2/\pi_F.$$

Using Eq. (A.3) of the Appendix, and assuming that b_i, S_i, and F_i have a jointly symmetric distribution, it is straightforward to show that

$$\text{plim}b_{\text{ols}} = \bar{\beta} + \lambda_0 + \psi_0\bar{S} = \bar{\beta} + \lambda_1 + \psi_1\bar{S} + (\lambda_2 + \psi_2\bar{S})\rho_{SF}^2/\pi_F,$$

[36] If F_i has an independent causal effect γ on earnings then Eq. (5″) includes a term γF_i. In this case the probability limit of the regression coefficient of F_i is $\gamma + \lambda_2$, and $\text{plim}b_{\text{iv}}$ includes a component equal to γ/π_F.

[37] Suppose that family background is measured by the average of mother's and father's education. The results in Table 2 suggest that $\pi_F \approx 0.4$ and $\rho_{SF}^2 \approx 0.3$, implying that the univariate OLS estimator will exceed the bivariate OLS by about $0.75\lambda_2$, while the IV estimator will exceed the bivariate OLS by $2.5\lambda_2$.

$$\text{plim} b_{\text{biv}} = \bar{\beta} + \lambda_1 + \psi_1 \bar{S},$$

$$\text{plim} b_{\text{iv}} = \bar{\beta} + \lambda_1 + \psi_1 \bar{S} + (\lambda_2 + \psi_2 \bar{S})/\pi_F.$$

Moreover, the probability limit of the coefficient on F_i in the bivariate regression is $\lambda_2 + \psi_2 \bar{S}$. In the presence of heterogeneity in b_i one can effectively reinterpret λ_1 as $(\lambda_1 + \psi_1 \bar{S})$ and λ_2 as $(\lambda_2 + \psi_2 \bar{S})$. Assuming that $\lambda_1 + \psi_1 \bar{S} \geq 0$, $\lambda_2 + \psi_2 \bar{S} \geq 0$, and $\pi_F > 0$, the probability limits of the three estimators continue to satisfy the inequalities

$$\bar{\beta} \leq \text{plim} b_{\text{biv}} \leq \text{plim} b_{\text{ols}} \leq \text{plim} b_{\text{iv}}.$$

In summary, unless $\lambda_1 = \lambda_2 = \psi_1 = \psi_2 = 0$ in the projection equations for the intercept and slope components of individual ability a_i and b_i, family background is *not* a legitimate instrument for schooling, even if family background has no direct causal effect on earnings. The addition of controls for family background may reduce the biases in the measured return to education, but may still lead to an upward-biased estimate of the average marginal return to schooling unless *all* of the unobserved ability components are absorbed by the family background controls (i.e., unless $\lambda_1 = \psi_1 = 0$). Finally, notice that in the special case where $\lambda_1 + \psi_1 \bar{S} = \lambda_2 + \psi_2 \bar{S}$, the upward bias in the estimated schooling coefficient from a bivariate model that controls for family background is equal to the probability limit of the coefficient on the family background variable itself. Under these circumstances, one can recover an unbiased estimate of the average marginal return to schooling by subtracting the family background coefficient from the own-schooling coefficient. This is equivalent to a "within-family" estimator, and will be discussed in more detail in the next section.

The preceding analysis assumes that true schooling is observable. In the more realistic case in which only a noisy measure of educational attainment is available, a comparison between the three estimators must take account of the differential impact of measurement errors on the univariate OLS, bivariate OLS, and IV estimators. Let R_0 represent the reliability of measured education and assume for the moment that F_i is measured without error. As noted earlier, the univariate OLS estimator is attenuated by the factor R_0:

$$\text{plim} b_{\text{ols}} = R_0[\bar{\beta} + \lambda_1 + \psi_1 \bar{S} + (\lambda_2 + \psi_2 \bar{S})\rho_{SF}^2/\pi_F].$$

The addition of F_i to the earnings model will tend to lead to greater attenuation of the coefficient on measured schooling, since some part of true education can be inferred from F_i. As shown in Appendix A, the bivariate OLS estimator is attenuated by a factor R_1:

$$\text{plim} b_{\text{biv}} = R_1[\bar{\beta} + \lambda_1 + \psi_1 \bar{S}],$$

where $R_1 = (R_0 - \rho_{SF}^2)/(1 - \rho_{SF}^2) < R_0$. For example, if $R_0 \approx 0.9$ and $\rho_{SF}^2 \approx 0.3$ then $R_1 \approx 0.85$. In contrast to either OLS estimator, the IV estimator is unaffected by measurement error. Thus, if F_i is measured without error, measurement errors in schooling will tend to *reinforce* the expected ranking of the univariate OLS, bivariate OLS, and IV

estimators by introducing the greatest attenuation bias in the bivariate OLS estimator, an intermediate bias in the univariate OLS estimator, and none in the IV estimator.

In many datasets family background information is collected from children or gathered retrospectively from older adults. In either case, one might expect F_i to contain substantial reporting errors. Indeed, Ashenfelter and Rouse (1998, Appendix 1) find that the reliability of twins' reports of their mother's education is about 80%, compared to a 90% reliability ratio for their own education. The presence of measurement errors in F_i creates a more complex expression for the probability limit of the bivariate OLS estimator. Specifically, the bivariate measurement error formula presented in the Appendix implies that

$$\text{plim}b_{\text{biv}} = R_1[\bar{\beta} + \lambda_1 + \psi_1\bar{S}] + (\lambda_2 + \psi_2\bar{S})(1 - R_F)\rho_{SF}^2/(\pi_F(1 - \rho_{SF}^2)),$$

where R_F is the reliability of measured family background. The second term in this expression is 0 if the true coefficient of family background in the bivariate model is 0 (i.e., if $\lambda_2 + \psi_2\bar{S} = 0$), or if $R_F = 1$. If the true coefficient of F_i is positive and $\pi_F > 0$, however, then measurement errors in F_i induce a positive bias in the schooling coefficient that may partially offset the direct attenuation effect of measurement error in S_i. For example, if $R_F \approx 0.8$, $\rho_{SF}^2 \approx 0.3$, and $\pi_F \approx 0.4$, the second term is on the order of 20% of the true coefficient of family background.

3.7. Models for siblings and twins

An alternative to the instrumental variables approach to the problem of causal inference is to study education and earnings outcomes for siblings, twins, or father–son/mother–daughter pairs. The key idea behind this strategy is that some of the unobserved differences that bias a cross-sectional comparison of education and earnings are reduced or eliminated within families.[38] For example, suppose that two observations (indexed by $j = 1$ or 2) are available for each family (indexed by i), and that the earnings of person j from family i are generated by

$$\log y_{ij} = a_0 + \bar{b}S_{ij} - 1/2k_1S_{ij} + a_{ij} + (b_{ij} - \bar{b})S_{ij}. \tag{15}$$

A "pure family effects" model is one in which $a_{ij} = a_i$ and $b_{ij} = b_i$. Consider the linear projections of a_i and $b_i - \bar{b}$ on the observed schooling outcomes of the two family members:

$$a_i = \lambda_1(S_{i1} - \bar{S}_1) + \lambda_2(S_{i2} - \bar{S}_2) + u_i, \tag{16a}$$

$$b_i - \bar{b} = \psi_1(S_{i1} - \bar{S}_1) + \psi_2(S_{i2} - \bar{S}_2) + v_i. \tag{16b}$$

Assuming that b_i, S_{i1}, and S_{i2} have a jointly symmetric distribution, Eq. (A.3) in Appendix A implies that the observed earnings outcomes of the family members are related to their

[38] Of course a within-family estimator can be given an IV interpretation: the instrument for schooling is the deviation of an individual's schooling from the average for his or her family.

schooling levels by

$$\log y_{i1} = c_1 + (\bar{\beta} + \lambda_1 + \psi_1 \bar{S}_1) S_{i1} + (\lambda_2 + \psi_2 \bar{S}_1) S_{i2} + e_{i1}, \tag{17a}$$

$$\log y_{i2} = c_2 + (\lambda_1 + \psi_1 \bar{S}_2) S_{i1} + (\bar{\beta} + \lambda_2 + \psi_2 \bar{S}_1) S_{i2} + e_{i1}, \tag{17b}$$

where c_1 and c_2 are constants and the residuals e_{ij} are orthogonal to both S_{i1} and S_{i2}. Eqs. (17a) and (17b) constitute a system of seemingly unrelated regressions.[39] Since there are no exclusion restrictions, the system can be estimated efficiently by applying OLS one equation at a time. Alternatively, one can construct the within-family difference in log earnings $\Delta \log y_i \equiv \log y_{i1} - \log y_{i2}$, and consider a model of the form

$$\Delta \log y_i = \mu_1 S_{i1} + \mu_2 S_{i2} + e_i. \tag{18}$$

Numerically, OLS estimates of the coefficients of (18) will be equal to the *differences* in the corresponding OLS estimates of the coefficients in (17a) and (17b).[40]

The attractiveness of the "pure family effects" model arises from the fact that one can potentially recover estimates of $\bar{\beta}$ from the differences in the coefficients of Eqs. (17), or from the coefficients of the differenced Eq. (18). For example, suppose there is no heterogeneity in b_i. In this case $\psi_1 = \psi_2 = 0$ in Eqs. (17a) and (17b), and therefore the coefficients of Eq. (18) satisfy

$$\text{plim}\mu_1 = \text{plim} - \mu_2 = \bar{\beta}.$$

A test of the hypothesis $\mu_1 = -\mu_2$ therefore provides a specification test of the "pure family effects" model when heterogeneity in the education slopes b_i is ignored.[41]

A "pure family effects" model is particularly plausible for identical twins, since identical twins share genetics and almost always share the same family background environment. For identical twins, it also seems natural to impose the symmetry conditions $\lambda_1 = \lambda_2 = \lambda$, $\psi_1 = \psi_2 = \psi$, and $\bar{S}_1 = \bar{S}_2 = \bar{S}$, since the identity of specific twins is arbitrary. With these simplifications Eqs. (17a) and (17b) reduce to

$$\begin{aligned} \log y_{i1} &= c_1 + (\bar{\beta} + \lambda + \psi \bar{S}) S_{i1} + (\lambda + \psi \bar{S}) S_{i2} + e_{i1} \\ &= c_1 + \bar{\beta} S_{i1} + (\lambda + \psi \bar{S})(S_{i1} + S_{i2}) + e_{i1}, \end{aligned} \tag{17a$'$}$$

[39] A system of equations like (17a) and (17b) is sometimes called a "correlated random effects" specification. The idea of projecting the unobservable residual component (i.e., the family effect) on the observed outcomes of the pair and then substituting the projection equation back into the earnings equation was popularized by Chamberlain (1982).

[40] If other covariates X_{ij} are included in the model then the first-differenced model has to contain X_{i1} and X_{i2} in order for the "adding up" condition to hold.

[41] Such tests have been widely used in other applications of the correlated random effects model: e.g., Jakubson (1988).

$$\log y_{i2} = c_2 + (\lambda + \psi\bar{S})S_{i1} + (\bar{\beta} + \lambda + \psi\bar{S})S_{i2} + e_{i2}$$

$$= c_2 + \bar{\beta}S_{i2} + (\lambda + \psi\bar{S})(S_{i1} + S_{i2}) + e_{i2}. \quad (17b')$$

These equations express log earnings of a particular twin in terms of his or her own education and the total (or average) education of the pair.[42] Under the assumptions of a "pure family effects" specification, all of the biases arising from the correlations between unobserved ability and schooling are loaded onto the coefficient associated with the total or average education of the family, and the own-schooling coefficient provides an unbiased estimate of the average marginal return to schooling. (This estimate is numerically equivalent to subtracting the estimated sibling education coefficient from the own schooling coefficient). Note that if the pure family effects and symmetry assumptions are satisfied, one can estimate $\bar{\beta}$ with data on earnings for only one twin, provided that both twin's schooling levels are known.[43]

In the case of siblings or father–son pairs it may be less plausible that individuals from the same family have exactly the same ability parameters. For example, older siblings may be treated differently than younger ones, leading to differences in their potential labor market outcomes.[44] The assumptions of a "pure family effects" model can be relaxed as follows. Consider the linear projections,

$$a_{i1} = \lambda_{11}(S_{i1} - \bar{S}_1) + \lambda_{12}(S_{i2} - \bar{S}_2) + u_{i1}, \quad (19a)$$

$$a_{i2} = \lambda_{21}(S_{i1} - \bar{S}_1) + \lambda_{22}(S_{i2} - \bar{S}_2) + u_{i1}, \quad (19b)$$

$$b_{i1} - \bar{b} = \psi_{11}(S_{i1} - \bar{S}_1) + \psi_{12}(S_{i2} - \bar{S}_2) + v_{i1}, \quad (19c)$$

$$b_{i2} - \bar{b} = \psi_{21}(S_{i1} - \bar{S}_1) + \psi_{22}(S_{i2} - \bar{S}_2) + v_{i2}, \quad (19d)$$

where u_{ij} and v_{ij} are orthogonal to S_{i1} and S_{i2}. For randomly-ordered siblings or fraternal twins it is natural to assume that the projection coefficients satisfy the symmetry restrictions: $\lambda_{11} = \lambda_{22}$, $\lambda_{12} = \lambda_{21}$, $\psi_{11} = \psi_{22}$, and $\psi_{12} = \psi_{21}$, although for father–son or mother–daughter pairs these assumptions are less appealing.[45] Substituting these equations into the earnings model (15) and considering the linear projection onto the observed schooling variables leads to a generalized version of Eqs. (17a) and (17b):[46]

$$\log y_{i1} = c_1 + \tau_{11}S_{i1} + \tau_{12}S_{i2} + e_{i1}, \quad (20a)$$

[42] Similar equations are derived by Ashenfelter and Rouse (1998).

[43] Exchangeability arguments suggest that symmetry should hold for a random ordering of twins in each family. However, if the "twin 1" sample is conditioned on employment and some of the individuals in the "twin 2" sample do not work, the ordering is no longer random, and symmetry might not be a valid restriction.

[44] See Kessler (1991). Kessler concludes that birth order has little or no effect on economic outcomes once family size is properly accounted for.

[45] For father–son pairs, Ashenfelter and Zimmerman (1997) propose a slightly generalized model in which $\lambda_{2j} = \alpha\lambda_{1j}$. They ignore heterogeneity in b_i.

$$\log y_{i2} = c_2 + \tau_{21} S_{i1} + \tau_{22} S_{i2} + e_{i2}, \tag{20b}$$

where

$$\tau_{11} = \bar{\beta} + \lambda_{11} + \psi_{11}\bar{S}_1, \qquad \tau_{12} = \lambda_{12} + \psi_{12}\bar{S}_1,$$

$$\tau_{21} = \lambda_{21} + \psi_{21}\bar{S}_2, \qquad \tau_{22} = \bar{\beta} + \lambda_{22} + \psi_{22}\bar{S}_2.$$

Clearly $\bar{\beta}$ is not identifiable from the seemingly unrelated regression coefficients in (20a) and (20b) even with the within-family symmetry assumptions, although if $\psi_{ij} = 0$ or $\bar{S}_1 = \bar{S}_2$ then symmetry imposes two linear restrictions on the coefficients ($\tau_{11} = \tau_{22}$ and $\tau_{21} = \tau_{12}$).

Nevertheless, it may be possible to place an upper bound on the average marginal return to schooling using data on fraternal twins or siblings. Specifically, suppose that $\lambda_{11} \geq \lambda_{12}$ and $\psi_{11} \geq \psi_{12}$; loosely, these assumptions mean that individual 1's own schooling is more informative about his or her ability than individual 2's schooling.[47] In this case,

$$\begin{aligned}\operatorname{plim}\tau_{11} - \tau_{12} &= \bar{\beta} + \lambda_{11} + \psi_{11}\bar{S}_1 - \lambda_{12} - \psi_{12}\bar{S}_1 \\ &= \bar{\beta} + (\lambda_{11} - \lambda_{12}) + (\psi_{11} - \psi_{12})\bar{S}_1 \\ &\geq \bar{\beta},\end{aligned}$$

so an *upper bound* estimator of $\bar{\beta}$ is $\tau_{11} - \tau_{12}$, the difference between the own-schooling effect and the other-family-member's-schooling effect in an equation for one family member's earnings.[48] Mechanically, this difference is equal to the coefficient of own-schooling when average family schooling is included in the regression, as in Eq. (17a′).[49]

Unfortunately, there is no guarantee that this bound is tighter than the bound implied by the cross-sectional OLS estimator. In other words, it is possible that the OLS estimator has a *smaller* upward bias than the within family estimator based on Eq. (17a). A necessary and sufficient condition for the within-family estimator to have a smaller asymptotic bias is

$$|\lambda_0 + \psi_0\bar{S}_1| > |\lambda_{11} - \lambda_{12} + (\psi_{11} - \psi_{12})\bar{S}_1|,$$

[46] Note that I am continuing to assume that (b_{ij}, S_{i1}, S_{i2}) have a jointly symmetric distribution.

[47] Assumptions on the relative magnitudes of the projection coefficients are most natural if S_{i1} and S_{i2} have the same variances. In that case, $\lambda_{11} - \lambda_{12} = A(\operatorname{cov}[a_{i1}, S_{i1}] - \operatorname{cov}[a_{i1}, S_{i2}])$ for some positive coefficient A; a similar expression holds for $\psi_{11} - \psi_{12}$.

[48] If the "pure family effects" and symmetry assumptions are satisfied then $\operatorname{plim}(\tau_{11} - \tau_{12}) = \hat{\beta}$.

[49] It is also closely related to the coefficient of the difference in schooling in an inter-family differenced model: $\Delta \log y_i = \tau_\Delta S_i + \Delta e_i$. This specification is appropriate if the symmetry restrictions $\lambda_{11} = \lambda_{22}$, $\lambda_{12} = \lambda_{21}$, $\psi_{11} = \psi_{22}$, $\psi_{12} = \psi_{21}$, and $\bar{S}_1 = \bar{S}_2$ are valid, in which case $\tau_{11} = \tau_{22}$ and $\tau_{21} = \tau_{12}$. For example, in the case of same-sex fraternal twins the identity of the individual twins is arbitrary so an "exchangeability" argument suggests that symmetry should hold. Under this assumption $\operatorname{plim}\tau_\Delta = \operatorname{plim}(\tau_{11} - \tau_{21}) = \operatorname{plim}(\tau_{11} - \tau_{12})$, although the estimate of τ_Δ is not mechanically equal to the difference in the estimates of τ_{11} and τ_{12}.

where λ_0 and ψ_0 are the projection coefficients defined in Eqs. (6a) and (6b). To illustrate the issues underlying the comparison between the OLS and within-family estimators, ignore heterogeneity in the earnings function intercepts a_{ij}, so that the relative asymptotic biases of the OLS and within-family estimators depend on the comparison between ψ_0 and $\psi_{11} - \psi_{12}$. Suppose first that the marginal costs of schooling are identical for members of the same family ($r_{ij} = r_i$) but that ability has no family component (i.e., $\text{cov}[b_{i1}, b_{i2}] = 0$). In this case *all* of the schooling differences within families are due to differences in ability, whereas across the population as a whole only a fraction $f = \sigma_b^2/(\sigma_b^2 + \sigma_r^2)$ of the variance of schooling is attributable to ability. As noted earlier, the endogeneity bias component in the cross-sectional OLS estimator is $\psi_0 = kf$. Using Eq. (19) it is easy to show that $\psi_{11} = kf/(1 - (1 - f)^2)$ and $\psi_{12} = -\text{kf}(1 - \text{f})/(1 - (1 - \text{f})^2)$. Hence $\psi_{11} - \psi_{12} = k$, implying that the within-family estimator has a greater endogeneity bias than the cross-sectional estimator.

At the other extreme, suppose that abilities are the same for members of the same family ($b_{ij} = b_i$) but that tastes are uncorrelated within families. In this case schooling differences within families are due entirely to differences in tastes, even though in the population as a whole a fraction f of the variance in schooling is due to differences in ability. Hence the within-family estimator is free of endogeneity biases whereas the OLS estimator has an endogeneity bias component $\psi_0 = kf$. More generally, the relative magnitudes of the endogeneity biases in the within-family and cross-sectional estimators depend on the relative contributions of ability differentials to the within-family and cross-sectional variances of schooling outcomes.[50] A within-family estimator will have a smaller bias if and only if ability differences are less important determinants of schooling within families than across the population as a whole.

Measurement error concerns play a fairly important role in the interpretation of estimates from sibling and family models. This is especially true in studies of identical twins, who tend to have very highly correlated education outcomes. For example, consider the estimation of Eq. (17a) using noisy measures of schooling for both twins. The multivariate measurement error formula implies that the probability limit of the coefficient on own-schooling is

$$\bar{\beta}\,\frac{R_0 - \rho^2}{1 - \rho^2} + (\lambda + \psi\bar{S})\frac{1 - R_0}{1 - \rho^2} \times \frac{\text{cov}[S_{i1}, S_{i1} + S_{i2}]}{\text{var}[S_{i1}]},$$

where R_0 is the reliability of measured schooling and ρ is the correlation of twin's schooling. Assuming that $R_0 \approx 0.9$ and $\rho \approx 0.75$ (see e.g., Ashenfelter and Rouse, 1998), this formula implies that the probability limit of the own schooling coefficient is roughly $0.8\bar{\beta} + 0.3(\lambda + \psi\bar{S})$.

Much of the twins literature focusses on estimation of a within-family differences model:

[50] A similar argument applies to the asymptotic biases in the two estimators associated with the correlation between a_{ij} and S_{ij}.

$$\Delta \log y_i = \tau_\Delta \Delta S_i + \Delta e_i.$$

Assuming that the "pure family effects" assumptions are satisfied and ignoring measurement error,

$$\text{plim} \tau_\Delta = \bar{\beta},$$

as can be seen by differencing Eqs. (20a) and (20b). The within-family differenced estimator is particularly susceptible to measurement error, however, since differencing within families removes much of the true signal in education. In particular, if the reliability of observed schooling is R_0 and the correlation between family members' schooling is ρ then the reliability of the observed difference in schooling is

$$R_\Delta = \frac{R_0(1-\rho)}{1-\rho R_0}.$$

When $R_0 \approx 0.9$ and $\rho \approx 0.75$, for example, $R_\Delta \approx 0.7$, implying a 30% attenuation bias in the OLS estimate of τ_Δ for identical twins. Among fraternal twins the correlation of schooling is lower: Ashenfelter and Krueger (1994) and Isacsson (1997) both estimate a correlation for fraternal twins of about 0.55. Assuming $R_0 \approx 0.9$ and $\rho \approx 0.55$, $R_\Delta \approx 0.8$, so one would expect a 20% attenuation bias in the OLS estimate of τ_Δ for fraternal twins.

3.8. Summary

Table 3 summarizes some of the key models, assumptions, and estimating equations that are useful in interpreting the returns to schooling literature. One estimation strategy not included in the table is instrumental variables based on a comparison between a quasi-experimental treatment group and a comparison group when the treatment has potentially different effects on the schooling attainment of different subgroups of the population. As noted above, under ideal conditions such an estimator will recover a weighted average of the marginal returns to education for different subgroups, where the weight applied to each subgroup is the change in schooling induced by the treatment. This weighted average may be above or below the average marginal return to education, depending on the nature of the intervention and the extent of heterogeneity in marginal returns.

Among the implications of the results summarized in Table 3 are:

1. The OLS estimator has two ability biases relative to the average marginal return to education ($\bar{\beta}$): one attributable to the correlation between schooling and the intercept of the earnings function (a_i), the other attributable to the correlation between schooling and the slope of the earnings function (b_i). The latter is unambiguously positive, but may be small in magnitude if the heterogeneity in returns to education is small (or if people lack perfect foresight about their abilities).
2. The necessary conditions for IV or control function estimators to yield a consistent estimate of $\bar{\beta}$ in the presence of heterogeneity in the returns to education are fairly strict. Plausible sources of exogenous variation in education choices (such as shifts in

Table 3
Summary of models, estimation methods, and probability limits of estimators

Model	Additional assumptions	Estimating equation	Probability limit of estimator
I. Ordinary least squares			
$y_i = a_0 + \bar{b}S_i - 1/2k_1S_i^2 + a_i + (b_i - \bar{b})S_i$ $a_i = \lambda_0(S_i - \bar{S}) + u_i$ $b_i = \bar{b} + \psi_0(S_i - \bar{S}) + v_i$	(a_i, b_i, S_i) jointly symmetric	$y_i = c + b_{ols}S_i$	b_{ols}: $\tilde{\beta} + \lambda_0 + \psi_0\bar{S}$, $\tilde{\beta} = \bar{b} - k_1\bar{S}$ with measurement error b_{ols}: $R_0\{\tilde{\beta} + \lambda_0 + \psi_0\bar{S}\}$, R_0 = reliability of S_i
IIa. Instrumental variables			
$y_i = a_0 + \bar{b}S_i - 1/2k_1S_i^2 + a_i + (b_i - \bar{b})S_i$ $S_i = \pi Z_i + \xi_i$	(a) $E[\eta_i\|Z_i] = 0$ (b) $E[a_i\|Z_i] = 0$ (c) $E[(b_i - \bar{b})\|Z_i] = 0$ (d) $E[(b_i - \bar{b})^2\|Z_i] = \sigma_b^2$ (e) $E[\xi_i\|b_i] = \rho_1(b_i - \bar{b})$	$S_i = \pi Z_i + \xi_i$ $y_i = c + b_{iv}\hat{S}_i$	b_{iv}: $\bar{\beta}$
IIb. Control function			
Same as IIa	(a)–(c) above plus (d) $E[a_i\|S_i, Z_i] = \lambda_1 S_i + \lambda_2' Z_i$ (e) $E[b_i\|S_i, Z_i] = \psi_1 S_i + \psi_2' Z_i$	$S_i = \pi Z_i + \xi_i$ $y_i = c + b_{cf}S_i + e_0\hat{\xi}_i + e_1S_i\hat{\xi}_i$	b_{cf}: $\bar{\beta}$

III. Family background models

$y_i = a_0 + \bar{b}S_i - 1/2k_1S_i^2$	(a_i, b_i, S_i, F_i) jointly	$y_i = c + b_{\text{biv}}S_i + gF_i$	b_{biv}: $\bar{\beta} + \lambda_1 + \psi_1\bar{S}$
$+a_i + (b_i - \bar{b})S_i$	symmetric	OR	g: $\lambda_2 + \psi_2\bar{S}$
$a_i = \lambda_1(S_i - \bar{S}) +$		$S_i = \pi_F F_i + e_i$	b_{iv}: $\bar{\beta} + \lambda_1 + \psi_1\bar{S} + (\lambda_2 + \psi_2\bar{S})/\pi_F$
$\lambda_2(F_i - \bar{F}) + u_i$			
$b_i = \bar{b} + \psi_1(S_i - \bar{S}) +$		$y_i = c + b_{\text{iv}}\check{S}_i$	with measurement error
$\psi_2(S_i - \bar{S}) + v_i$			
$\rho =$ correlation (S_i, F_i)			b_{biv}: $R_1\{\bar{\beta} + \lambda_1 + \psi_1\bar{S}\}$, $R_1 = (R_0 - \rho)/(1 - \rho)$

IV. Sibling/twin models

$y_{ij} = a_0 + \bar{b}S_{ij} - 1/2k_1S_{ij}^2$	$(a_{ij}, b_{ij}, S_{i1}, S_{i2})$ jointly	$y_{i1} = c_1 + \tau_{11}S_{i1} + \tau_{12}S_{i2}$	τ_{11}: $\bar{\beta} + \lambda_{11} + \psi_{11}\bar{S}_1$, τ_{12} : $\lambda_{12} + \psi_{12}\bar{S}_1$
$+a_{ij} + (b_{ij} - \bar{b})S_{ij}$	symmetric	$y_{i2} = c_2 + \tau_{21}S_{i1} + \tau_{22}S_{i2}$	τ_{22}: $\bar{\beta} + \lambda_{22} + \psi_{22}\bar{S}_2$, τ_{21}: $\lambda_{21} + \psi_{21}\bar{S}_2$
$a_{ij} = \lambda_j 1(S_{i1} - \bar{S}_1)$			
$+\lambda_{j2}(S_{i2} - \bar{S}_2) + u_{ij}$			
$b_{ij} = \bar{b} + \psi_j 1(S_{i1} - \bar{S}_1)$	above plus $\lambda_{12} = \lambda_{21}$;	as above	$\tau_{11} = \tau_{22}$: $\bar{\beta} + \lambda_{11} + \psi_{11}\bar{S}$
$+\psi_{j2}(S_{i2} - \bar{S}_2) + v_{ij}$	$\psi_{12} = \psi_{21}$;		
	$\lambda_{11} = \lambda_{22}$; $\psi_{11} = \psi_{22}$;	OR	$\tau_{12} = \tau_{21}$: $\lambda_{12} + \psi_{12}\bar{S}$
	$\bar{S}_1 = \bar{S}_2 = \bar{S}$		
	$j = 1, 2$ "exchangeable"	$\Delta y_i = \tau_\Delta \Delta S_i$	τ_Δ: $\bar{\beta} + \lambda_{11} - \lambda_{12} + (\psi_{11}\text{-}\psi_{12})\bar{S}$
			with measurement error
			τ_Δ: $R_\Delta\{\bar{\beta} + \lambda_{11} - \lambda_{12} + (\psi_{11} - \psi_{12})\bar{S}\}$
			R_Δ = reliability of ΔS_i

the cost of schooling) may not satisfy these conditions, in which case IV will recover a weighted average of marginal returns for the affected subgroups.

3. If the OLS estimator is upward-biased by unobserved ability, one would expect an IV estimator based on family background to be even more upward-biased.
4. If twins or siblings have *identical* abilities (and the distributions of abilities among twins are the same as those in the population as a whole) then a within-family estimator will recover an asymptotically unbiased estimate of the average marginal return to education. Otherwise, a within-family estimator may be more or less biased by unobserved ability effects than the corresponding cross-sectional OLS estimator, depending on the relative fraction of the variance in schooling attributable to ability differences within families versus across the population.
5. Measurement error biases are potentially important in interpreting the estimates from different procedures. Conventional OLS estimates are probably downward-biased by about 10%; OLS estimates that control for family background (or the education of a sibling) may be downward-biased by 15% or more; and within-family differenced estimates may be downward-biased by 20–30%, with the upper range more likely for identical twins.

4. A selective review of recent empirical studies

I now turn to a selective review of the recent literature on estimating the return to schooling. I summarize three sets of findings: instrumental variables estimates of the return to education based on institutional features of the education system; estimates based on either controlling for family background or using family background as an instrument for schooling; and estimates based on the schooling and earnings of twins. I also briefly review recent efforts to model observable heterogeneity in the returns to schooling. One strand of literature that I do not consider are studies of the return to schooling that attempt to control for ability using observed test scores. Some of the subtle issues involved in developing a causal framework for the interpretation of test scores, schooling outcomes, and earnings are considered in Griliches (1977, 1979), Chamberlain (1977) and Chamberlain and Griliches (1975, 1977).

4.1. Instrumental variables based on institutional features of the school system

One of the most important new directions of research in the recent literature on schooling is the use of institutional features of the schooling system as a source of credible identifying information for disentangling the causal effects of schooling.[51] Table 4 summarizes seven recent studies that estimate the return to schooling using instrumental variables

[51] This idea is also proving useful in studies of the effect of school quality. For example, Angrist and Lavy (1997) use information on maximum class size to identify the effect of class size on student achievement.

Table 4
OLS and IV estimates of the return to education with instruments based on features of the school system[a]

Author	Sample and instrument		Schooling coefficients	
			OLS	IV
1. Angrist and Krueger (1991)	1970 and 1980 Census Data, Men. Instruments are quarter of birth interacted with year of birth. Controls include quadratic in age and indicators for race, marital status, urban residence	1920–1929 cohort in 1970	0.070 (0.000)	0.101 (0.033)
		1930–1939 cohort in 1980	0.063 (0.000)	0.060 (0.030)
		1940–1949 cohort in 1980	0.052 (0.000)	0.078 (0.030)
2. Staiger and Stock (1997)	1980 Census, Men. Instruments are quarter of birth interacted with state and year of birth. Controls are same as in Angrist and Krueger, plus indicators for state of birth. LIML estimates	1930–1939 cohort in 1980	0.063 (0.000)	0.098 (0.015)
		1940–1949 cohort in 1980	0.052 (0.000)	0.088 (0.018)
3. Kane and Rouse (1993)	NLS Class of 1972, Women. Instruments are tuition at 2 and 4-year state colleges and distance to nearest college. Controls include race, part-time status, experience. Note: Schooling measured in units of college credit equivalents	Models without test scores or parental education	0.080 (0.005)	0.091 (0.033)
		Models with test scores and parental education	0.063 (0.005)	0.094 (0.042)
4. Card (1995b)	NLS Young Men (1966 Cohort) Instrument is an indicator for a nearby 4-year college in 1966, or the interaction of this with	Models that use college proximity as instrument (1976 earnings)	0.073 (0.004)	0.132 (0.049)

Table 4 (*continued*)

Author	Sample and instrument		Schooling coefficients	
			OLS	IV
	parental education. Controls include race, experience (treated as endogenous), region, and parental education	Models that use college proximity × family background as instrument	–	0.097 (0.048)
5. Conneely and Uusitalo (1997)	Finnish men who served in the army in 1982, and were working full time in civilian jobs in 1994. Administrative earnings and education data. Instrument is living in university town in 1980. Controls include quadratic in experience and parental education and earnings.	Models that exclude parental education and earnings	0.085 (0.001)	0.110 (0.024)
		Models that include parental education and earnings	0.083 (0.001)	0.098 (0.035)
6. Maluccio (1997)	Bicol Multipurpose Survey (rural Philippines): male and female wage earners age 20–44 in 1994, whose families were interviewed in 1978. Instruments are distance to nearest high school and indicator for local private high school. Controls include quadratic in age.	Models that do not control for selection of employment status or location	0.073 (0.011)	0.145 (0.041)
		Models with selection correction for location and employment	0.063 (0.006)	0.113 (0.033)
7. Harmon and Walker (1995)	British Family Expenditure Survey 1978–1986 (men). Instruments are indicators for changes in the minimum school leaving age in 1947 and 1973. Controls include quadratic in age, year, survey and region and region		0.061 (0.001)	0.153 (0.015)

[a] Notes: see text for sources and information on individual studies.

based on this idea. For each study I report both OLS and IV estimates derived from the same sample with the same control variables.

Angrist and Krueger's (1991) landmark study uses an individual's quarter of birth (interacted with year of birth or state of birth in some specifications) as an instrument for schooling. They show that men born from 1930 to 1959 with birth dates earlier in the year have slightly less schooling than men born later in the year – an effect they attribute to compulsory schooling laws. Angrist and Krueger note that people born in the same calendar year typically start school at the same time. As a result, individuals born earlier in the year reach the minimum school-leaving age at a lower grade than people born later in the year, allowing those who want to drop out as soon as legally possible to leave school with less education. Assuming that quarter of birth is independent of taste and ability factors, this phenomenon generates exogenous variation in education that can be used in an IV estimation scheme. It is worth emphasizing that compulsory schooling laws presumably raise the education of people who would otherwise choose low levels of schooling. If these individuals have higher or lower marginal returns to education than other people, a quarter-of-birth-based IV estimator may over- or under-estimate the average marginal return to education in the population as a whole.

Angrist and Krueger's empirical analysis confirms that the quarterly pattern in school attainment is paralleled by a similar pattern in earnings. As shown in Table 4, their IV estimates of the return to education are typically higher than the corresponding OLS estimates, although for some cohorts and specifications the two estimators are very close, and in no case is the difference between the IV and OLS estimators statistically significant.

Angrist and Krueger's findings have attracted much interest and some criticism. Bound et al. (1995) point out that several of Angrist and Krueger's IV models (specifically, those that use interactions between quarter of birth and state of birth as predictors for education) include large numbers of weak instruments, and are therefore asymptotically biased toward the corresponding OLS estimates. This "weak instruments" bias is less of an issue for the specifications reported in Table 4, which rely on a more parsimonious set of instruments. Moreover, to the extent that Angrist and Krueger's IV estimates are *above* the corresponding OLS estimates, one might infer that asymptotically unbiased estimates of the causal effect of education are even higher. This is confirmed by the findings of Staiger and Stock (1997), who re-analyze the 1980 Census samples used by Angrist and Krueger and compute a variety of asymptotically valid confidence intervals for standard IV and limited information maximum likelihood (LIML) estimates. Staiger and Stock's preferred LIML estimates, utilizing quarter of birth interacted with state of birth and year of birth as instruments, are reported in row 2 of Table 4. These are somewhat above the corresponding conventional IV estimates and, 50–70% higher than the OLS estimates.

A second criticism of Angrist and Krueger's findings, raised by Bound and Jaeger (1996), is that quarter of birth may be correlated with unobserved ability differences. Bound and Jaeger examine the schooling outcomes of earlier cohorts of men who were not subject to compulsory schooling institutions and find some evidence of seasonal

patterns. They also discuss evidence from the sociobiology and psychobiology literature which suggests that season of birth is related to family background and the incidence of mental illness.

To evaluate the differences in family background by quarter of birth for cohorts roughly comparable to the ones in Angrist and Krueger's study, I compared the mean levels of parents' education by quarter of birth for children under 1 year of age in the 1940 Census.[52] The mean years of education for mothers of children born in quarters I, II, III, and IV, are 9.04, 8.95, 8.97, and 8.95, respectively (with standard errors of about 0.05). The corresponding means of father's education are 8.61, 8.50, 8.52, and 8.58. These comparisons give no indication that children born in the first quarter come from relatively disadvantaged family backgrounds, and suggest that the seasonality patterns identified by Angrist and Krueger are probably not caused by differences in family background.

The third study summarized in Table 4, by Kane and Rouse (1993), is primarily concerned with the relative labor market valuation of credits from regular (4-year) and junior (2-year) colleges. Their findings suggest that credits awarded by the two types of colleges are interchangeable: in light of this conclusion they measure schooling in terms of total college credit equivalents. In analyzing the earnings effects of college credits, Kane and Rouse compare OLS specifications against IV models that use the distance to the nearest 2-year and 4-year colleges and state-specific tuition rates as instruments. Their IV estimates based on these instruments are 15–50% above the corresponding OLS specifications.

Two subsequent studies by Card (1995b) and Conneely and Uusitalo (1997) examine the schooling and earnings differentials associated with growing up near a college or university. The Card (1995b) study finds that when college proximity is used as an instrument for schooling in the National Longitudinal Survey (NLS) Young Men sample, the resulting IV estimator is substantially above the corresponding OLS estimator, although rather imprecise. Consistent with the idea that accessibility matters more for individuals on the margin of continuing their education, college proximity is found to have a bigger effect for children of less-educated parents. This suggests an alternative specification that uses interactions of college proximity with family background variables as instruments for schooling, and includes college proximity as a direct control variable. The IV estimate from this interacted specification is somewhat lower than the estimate using college proximity alone, but still about 30% above the OLS estimate.

The Conneely and Uusitalo (1997) study utilizes a very rich Finnish dataset that combines family background information, military test scores, and administrative earnings data for men who served in the army in 1982. Like Kane and Rouse (1993) and Card (1995b) they find that IV estimates of the returns to schooling based on college proximity exceed the corresponding OLS estimates by 20–30%, depending on what other controls are added to the model. It is worth noting that all three of these studies report models that

[52] Quarter of birth is only reported in the 1940 Census for children under 1 year of age. There are 19,089 children under 1 year of age in the public use file, of whom 98.4% can be matched to a female head of household and 95.3% can be matched to a male head of household.

control for a fairly detailed set of family background characteristics. Such controls are desirable if families that live near colleges have different family backgrounds, and if family background has some independent causal effect on earnings. Conneely and Uusitalo's IV estimate controlling for parental education and earnings is below the IV estimate that excludes these controls, but is still above the simplest OLS estimate without family background controls. Despite the rather large size of their sample (about 22,000 observations) and the very high quality of their underlying data, Conneely and Uusitalo's IV estimates are somewhat imprecise, and are not significantly different from their OLS estimates.[53]

The sixth study in Table 4, by Maluccio (1997), applies the school proximity idea to data from the rural Philippines. Maluccio combines education and earnings information for a sample of young adults with data for their parents' households, including the distance to the nearest high school and an indicator for the presence of a local private high school. These variables have a relatively strong effect on completed education in this sample. Maluccio estimates OLS and conventional IV models using school proximity as an instrument, as well as IV models that include a selectivity correction for employment status and location. Both IV estimates are substantially above the corresponding OLS estimates. Maluccio's analysis suggests that the reliability of his schooling variable is somewhat lower than in conventional US or European datasets ($R_0 \approx 0.8$), accounting for some of the gap between the IV and OLS estimates. Unfortunately, Maluccio does not present OLS or IV models that control for family background. Rather, he presents IV models that use parental education and wealth as additional instruments for education, leading to slightly smaller but somewhat more precise IV estimates.

The final study summarized in Table 4, by Harmon and Walker (1995), examines the returns to education among a relatively large sample of British male household heads. Harmon and Walker use as instrumental variables for schooling a pair of dummy variables that index changes in the minimum school leaving age in Britain – from 14 to 15 in 1947, and from 15 to 16 in 1973. These are effectively cohort dummies that distinguish between men born before 1932, those born from 1933 to 1957, and those born after 1957. As shown in Table 4 their IV estimate is considerably above their OLS estimate (2.5 times higher) and is relatively precise. There are several aspects of their estimation strategy that suggest the need for caution in the interpretation of these findings, however. Most importantly, the 1947 law change – which is the major source of identification in their results – came just after World War II.[54] Moreover, Harmon and Walker do not allow for systematic growth in educational attainment for consecutive cohorts of men, other than that attributable to the law changes in 1947 and 1973.[55] Both these factors may bias their IV estimator up.

[53] Conneely and Uusitalo also implement a more general control function estimator, as described above.

[54] Ichino and Winter-Ebmer (1998) document that across Europe, the educational attainments of children born between 1930 and 1935 were substantially below those of children born just earlier or later.

[55] Their specifications control for age and survey year. One can infer the presence of important cohort effects from the fact that their survey year effects show a 0.5 year rise in educational attainment between surveys in 1979 and 1986, controlling for age and the school leaving age indicators.

In addition to the studies included in Table 4, a number of other recent studies have used IV techniques to estimate the return to schooling. One innovative example is Hausman and Taylor (1981), which uses the means of three time-varying covariates (age and indicators for the incidence of bad health and unemployment) as instruments for education in a panel data model of earnings outcomes for prime-age men. Hausman and Taylor find that the return to schooling rises from about 0.07 in OLS specifications to 0.12–0.13 in their IV specifications. Although more recent studies have not directly followed Hausman and Taylor's methodology, their use of mean age as an instrument for schooling is equivalent to using a linear cohort variable, and is thus similar in spirit to Harmon and Walker.

A very recent study by Ichino and Winter-Ebmer (1998) also utilizes birth cohort as a source of variation in schooling outcomes. In particular, Ichino and Winter-Ebmer focus on the earnings and schooling outcomes of Austrian and German men born from 1930 to 1935. They argue that World War II had a particularly strong effect on the educational attainment of children who reached their early teens during the war and lived in countries directly subject to hostilities. Using data for 14 countries they find relatively big differences in completed education for children in the 1930–1935 cohort in countries that were most heavily affected by the war (e.g., Germany, Austria and the UK) but relatively small differences for this cohort in other places (e.g., the US and Ireland). When they use an indicator for the 1930–1935 cohort as an instrument for low educational attainment they find that the earnings disadvantage roughly *doubles* from its OLS value. While one might be concerned that the 1930–1935 cohort suffered other disadvantages besides their disrupted education careers, these results are comparable to Harmon and Walker's (1995) in terms of the magnitude of the IV/OLS gap.

Another study not reported in Table 4, by Angrist and Krueger (1992), examines the potential effect of "draft avoidance" behavior on the education and earnings of men who were at risk of induction in the 1970–1973 Vietnam war draft lotteries. Since enrolled students could obtain draft exemptions, many observers have argued that the draft lottery led to higher college enrollment rates, particularly for men whose lottery numbers implied the highest risk of induction. If true, one could use draft lottery numbers – which were randomly assigned by day of birth – as instruments for education. While Angrist and Krueger (1992) report IV estimates based on this idea, subsequent research (Angrist and Krueger, 1995) showed that the link between lottery numbers and completed education is quite weak. In fact, the differences in education across groups of men with different lottery numbers are not statistically significant. Thus, the IV estimates are subject to the weak instruments critique of Bound et al. (1995), and are essentially uninformative about the causal effect of education.[56]

A conclusion that emerges from the results in Table 4 and from other IV-based studies is that instrumental variables estimates of the return to schooling typically exceed the corresponding OLS estimates – often by 30% or more. If one assumes on a priori grounds that

[56] The conventional IV estimates are typically equal to or just above the OLS estimates. Angrist and Krueger (1995) propose a "split sample" IV method to deal with the weak instruments problem. The split-sample IV estimates are all very imprecise.

OLS methods lead to upward-biased estimates of the true return to education, the even larger IV estimates obtained in many recent studies present something of a puzzle. A number of hypotheses have been offered to explain this puzzle. The first – suggested by Bound and Jaeger (1996), for example – is that the IV estimates are *even further* upward biased than the corresponding OLS estimates by unobserved differences between the characteristics of the treatment and comparison groups implicit in the IV scheme. This is certainly a plausible explanation for some part of the gap between OLS and IV in studies that do not control directly for family background, but it is less compelling for studies that include family background controls.

A second explanation – proposed by Griliches (1977) and echoed by Angrist and Krueger (1991) – is that ability biases in the OLS estimates of the return to schooling are relatively small, and that the gaps between the IV and OLS estimates in Table 4 reflect the *downward* bias in the OLS estimates attributable to measurement errors. The imprecision of most of the IV estimates in Table 4 makes it difficult to rule out this explanation on a study-by-study basis. Since measurement error bias by itself can only explain a 10% gap between OLS and IV, however, it seems unlikely that so many studies would find large positive gaps between their IV and OLS estimates simply because of measurement error.[57]

A third possibility, suggested in a recent overview of the returns to education literature by Ashenfelter and Harmon (1998), is "publication bias". They hypothesize that in searching across alternative specifications for a statistically significant IV estimate, a researcher is more likely to select a specification that yields a large point estimate of the return to education. As evidence of this behavior they point to a positive correlation across studies between the IV-OLS gap in the estimated return to education and the sampling error of the IV estimate.[58]

While all three of these explanations have some appeal, I believe a fourth explanation based on underlying heterogeneity in the returns to education is also potentially important. Factors like compulsory schooling or the accessibility of schools are most likely to affect the schooling choices of individuals who would otherwise have relatively low schooling. If these individuals have higher-than-average marginal returns to schooling, then instrumental variables estimators based on compulsory schooling or school proximity might be expected to yield estimated returns to schooling above the corresponding OLS estimates. A necessary condition for this phenomenon is that marginal rates of return to schooling are negatively correlated with the level of schooling across the population. In the model presented in Section 3, the covariance of the return to schooling with the level of schooling is $E[\beta_i(S_i - \bar{S})] = (kf - k_1)\text{Var}[S_i]$, where $k = k_1 + k_2$ and f is the fraction of the variance of schooling outcomes attributable to variation in ability. If individual discount rates are constant (i.e., $k_2 = 0$) this covariance is necessarily negative. Even if individuals have

[57] One caveat to this conclusion is the possibility that measurement errors are larger, or more systematically correlated with schooling levels, for individuals most affected by the interventions underlying the analyses in Table 4. Kane et al. (1997) find some evidence of this.

[58] Across the studies in Table 4 the IV-OLS gap is negatively related to the sampling error of the IV estimate, although the correlation is positive if the Harmon–Walker study is excluded.

increasing marginal discount rates (because of taste factors or financial constraints) marginal returns to education will be higher for less-educated individuals if ability differences are not "too important" in the determination of schooling outcomes, and if the marginal return to schooling is decreasing. In this case, IV estimates of the return to schooling based on institutional changes that raise schooling levels among less-educated subgroups may well exceed the corresponding OLS estimates.

4.2. Estimators using family background as a control or instrument

Table 5 summarizes some findings on the use of family background (typically parental education or the education of a sibling) as either a control variable or instrument in models of the return to education. For most of the studies presented in the table, I report three estimates of the return to education: an OLS estimate that excludes family background controls; an OLS estimate that controls for one or more family background characteristics; and an IV estimator that uses *the same* family background variable(s) as an instrument for education. For two of the studies (Miller et al., 1995; Ashenfelter and Rouse, 1998) I also present measurement-error corrected IV estimates for models that include both an individual's education and his or her sibling's education, using multiple reports of the siblings' education as instruments.[59] It should be noted that most of the studies described in Table 5 do not focus directly on the specifications I have summarized, but rather report these results incidentally.

The first group of studies in the table utilize parental education as a family background indicator. The Card (1995b) and Conneely and Uusitalo (1997) studies have already been described.[60] I prepared the estimates for the General Social Survey (GSS) sample specifically for this review.[61] The Ashenfelter and Zimmerman (1997) paper uses father's education as a background variable in one set of models, and brother's education in another. With the exception of the results for women in the GSS, the results for these four studies are remarkably consistent. In all four cases the addition of parental education as a control variable (or set of controls) lowers the measured return to education by 5–10% (about the magnitude of the decline expected on account of measurement error factors alone); while the use of parental education as an instrument leads to IV estimates that are at least 15% above the corresponding OLS estimates. Moreover (although not shown in the table), the coefficient of the parental education variable itself is positive and significant, but small in magnitude. For women in the GSS sample the addition of mother's education has essentially no effect on the return to a woman's own education, and higher mother's

[59] Specifically, following the lead of Ashenfelter and Krueger (1994), both Miller et al. and Ashenfelter and Rouse make use of information collected from twins on their own and their sibling's education.

[60] The IV estimate associated with the data in my 1995b study is not reported in the published version of the paper.

[61] The GSS has the advantage of including large samples of men and women. Earnings information in this survey pertains to annual income: I imputed interval midpoints to the categorical data in the survey.

Table 5
Estimates of the return to education with and without controlling for family background, and IV estimates using family background[a]

Author	Sample and family background variable(s)		OLS coefficients		IV coefficient
			No control	Control	
1. Card (1995b)	NLS Young Men (see Table 4). Family background variables are both parents' education (main effects and interactions) plus family structure[b]		0.073 (0.006)	0.069 (0.006)	0.084 (0.009)
2. This chapter	General Social Survey of adult household heads age 24–61, 1974–1996 data. Annual earnings (imputed from categorical data). Controls include cubic in age, race, survey year and region. Family background variable is mother's education	Men ($N = 7860$)	0.073 (0.003)	0.067 (0.003)	0.106 (0.007)
		Women ($N = 7500$)	0.112 (0.004)	0.113 (0.004)	0.110 (0.011)
3. Conneely and Uusitalo (1997)	Finnish male veterans (see Table 4). Family background variable is parent's education		0.085 (0.001)	0.082 (0.001)	0.114 (0.006)
4. Ashenfelter and Zimmerman (1997)	NLS Young Men (1966 Cohort) merged with NLS Older Men. Family background variables are brother's or father's education. Controls include quadratic in age	Brother 1, using other brother's education	0.059 (0.014)	0.052 (0.015)	0.080 (0.027)
		Sons, using father's education	0.057 (0.009)	0.049 (0.009)	0.109 (0.025)
5. Miller et al. (1995)	Australian Twins Register (male and female identical twins). Income imputed from occupation. Family background variable is twin's education. Controls include quadratic in age and marital status	No allowance for measurement error	0.064 (0.002)	0.048 (0.003)	–
		IV using twin's report[c]	0.073 (0.003)	0.078 (0.009)	–
6. Ashenfelter and Rouse (1998)	1991–1993 Princeton Twins Survey (men and women). Identical twins. Family background variable is twin's education. Controls include gender, race, and quadratic in age	No allowance for measurement error	0.102 (0.010)	0.092 –	
		IV using twin's report[c]	0.112 –	0.108 –	–

Table 5 (*continued*)

Author	Sample and family background variable(s)		OLS coefficients		IV coefficient
			No control	Control	
7. Isacsson (1997)	Swedish Twins Registry (men and women). Same-sex twins born 1926–1958. Administrative earnings data (average of 3 years). Family background variable is other twin's education.	Identical twins	0.046 (0.001)	0.040 –	0.055 (0.002)
		Fraternal twins	0.047 (0.001)	0.046 –	0.054 (0.002)

[a] Notes: See text for sources and information on individual studies.

[b] In this study the IV specification treats education and experience as endogenous and uses family background variables and age as instruments.

[c] In these specifications each twin's education is instrumented by the other twin's report of their education.

education has a very small negative effect on earnings. As a consequence the IV estimate for the GSS female sample is slightly lower than the OLS estimate.[62]

The fifth, sixth and seventh studies described in Table 5 all utilize samples of twins: in each case family background is measured by a sibling's education. Interestingly, the effect of adding a twin's education in these samples is similar to the effect of adding parental background in the other studies: the coefficient of own-schooling falls by 10–25%. Since twin's education levels are even more highly correlated than father–son or sibling education levels, the magnitude of this drop is not far off the decline attributable to measurement error factors alone. The Miller et al. and Ashenfelter–Rouse studies allow a direct test of the "pure measurement error" explanation, since in both cases the authors report estimates for IV models that include both twins' education levels (as reported by one twin) instrumented by the education levels reported by the other twin.[63] As shown in the table, the measurement-error corrected estimates of the return to own-education with controls for twin's-education are about equal to the corresponding measurement error-corrected OLS estimates that do not control for family background.

Based on these findings for twins, and the results in the other studies in Table 5, I conclude that whatever biases exist in conventional OLS estimates of the return to education are also present in models that control for family background. Apart from an effect attributable to measurement error, the return to education is about the same when controls are introduced for the education of one's parents or siblings. In the context of the models summarized in Table 3, this finding suggests that the bias component in the simple OLS

[62] Recall that these three estimators are mechanically linked. If mother's education has a negative effect controlling for daughter's education, then the IV estimate using mother's education as an instrument is necessarily below the OLS estimate.

[63] Ashenfelter and Rouse (1998) actually report estimates of models that include S_{i1} and $(S_{i1} + S_{i2})/2$ (i.e., average family education). These coefficients can be "unscrambled" to show the direct effects of S_{i1} and S_{i2}, although there is not enough information to construct standard errors for these effects.

estimator, $\lambda_0 + \psi_0\bar{S}$, is about the same size as, or only slightly bigger than, the bias in the estimator that controls for family background, $\lambda_1 + \psi_1\bar{S}$.

On the other hand, measures of family background such as parental or sibling education typically exert a small positive effect on earnings (i.e., the term $\lambda_2 + \psi_2\bar{S}$ is positive). Thus, IV estimates using family background as an exogenous determinant of schooling are often (but not always) substantially above the corresponding OLS estimates.[64] This conclusion is potentially important for interpreting other IV estimates of the return to education based on factors like proximity to college or other institutional features of the education system. To the extent that individuals in the treatment and control groups of a quasi-experimental analysis have different family backgrounds, one might expect a positive upward bias in the resulting IV estimators. The IV results in Table 5 suggest that it is particularly important to control for family background (or verify that family background is the same in the treatment and control groups) in any instrumental variables analysis of the return to schooling.

As noted in Section 3, although the addition of controls for family background will not necessarily lead to consistent estimates of the true return to schooling, under certain assumptions estimates from models that control for family background can be used to obtain consistent estimates of the average marginal return to education. Specifically, if one assumes that $\lambda_1 + \psi_1\bar{S} = \lambda_2 + \psi_2\bar{S}$, the upward bias in the estimated own schooling coefficient is equal to the probability limit of the family background variable's coefficient. Thus one can subtract the latter from the former and obtain a consistent estimate of $\bar{\beta}$. Given that family background variables like the education of a parent or sibling typically exert a small positive effect on earnings, application of this procedure to the studies in Table 5 would lead to estimates of the average marginal return to schooling that are somewhat below the OLS estimates. Subtraction of the coefficient of a parent's or sibling's education from the own schooling coefficient is equivalent to a within-family estimator. Since the assumptions required to justify this estimator are most appealing in the case of twins, I defer a more detailed discussion to the next section.

Under slightly weaker assumptions but with more information – specifically, with information on the earnings of the family member whose data is used as a control – it still may be possible to estimate the average marginal return to schooling. In particular, under the "pure family effects" assumption that siblings or parents share the same abilities, one can derive an estimate of $\bar{\beta}$ from the coefficients of a seemingly unrelated regression of each family members' earnings on his or her own schooling, and the other

[64] This conclusion differs slightly from Griliches' (1979, p. S59) tentative conclusion that "measured parental characteristics... appear to affect earnings primarily via their effect on the level of achieved schooling. The market does not appear to pay for them directly." Another dimension of family background that seems to have some effect on education of women is the sex composition of one's siblings. Butcher and Case (1994) show that women with brothers (rather than sisters) have slightly more education. They also use sex composition as an instrument for schooling and find much larger IV than OLS estimates of the return to schooling. Even though sex composition is random, it is unclear that its only effect on earnings is via education: thus Butcher and Case's (1994) IV estimates may be biased.

family members schooling (see Eqs. (17a) and (17b)). Ashenfelter and Zimmerman report estimates from this procedure applied to brothers and father–son pairs, with and without corrections for measurement error biases. Their estimation methods ignore heterogeneity in the returns to schooling. This is not a problem for their sample of brothers, who have roughly the same mean education, but may be more of an issue for their father–son sample, since the sons have about four years more education than the fathers.[65]

After accounting for plausible measurement error biases, Ashenfelter and Zimmerman's findings for brothers imply estimates of $\bar{\beta}$ about equal to the corresponding OLS estimates. Their estimates for father–son pairs are more sensitive to assumptions about whether the true return to education is the same for fathers and sons, and whether fathers and sons are exchangeable in the projection equations for the latent family ability term. Their least restrictive specifications suggest a slightly lower estimate of $\bar{\beta}$ for fathers than the corresponding OLS estimate, but a much lower estimate of $\bar{\beta}$ for sons. Given their results for brothers, however, an alternative interpretation is that the "pure family effects" assumption is inappropriate for father–son pairs. In fact, Ashenfelter and Zimmerman find that a slightly modified model that allows latent family ability to have a differential effect on the intercepts of fathers' and sons' earnings equations seems to fit the data fairly well. After correcting for measurement error, this specification implies estimates of $\bar{\beta}$ for fathers and sons that are 25–50% lower than the corresponding OLS estimates.

4.3. Studies of education and earnings using twins

Table 6 summarizes five recent studies that compare the education and earnings of twins. Two features of these studies contrast with the earlier literature on twins surveyed by Griliches (1979). First, the samples in the recent literature are relatively large, and tend to include a broader range of age and family background groups. Second, following the lead of Ashenfelter and Krueger's (1994) innovative paper, most of the recent studies squarely address the problem of measurement error. For each study I report a cross-sectional (OLS) return to education, and two within-family differenced estimates: one estimated by OLS and the other corrected for measurement error.

The Ashenfelter and Rouse (1998) study utilizes 3 years of data collected in the Princeton Twins Survey (PTS): their sample includes 340 pairs of identical twins, 60% of whom are women. As shown for the two specifications in Table 6, Ashenfelter and Rouse's within-family estimates of the return to education are about 30% lower than their corresponding OLS estimates. This finding contrasts with the results in Ashenfelter and Krueger (1994) based on only 1 year of data from the PTS, which indicated a *bigger* within-family than OLS estimate.[66] The PTS questionnaire asked each twin their own education and their sibling's education. This extra set of responses allow Ashenfelter and

[65] If there is heterogeneity in returns to education Eqs. (17a) and (17b) imply that the coefficients of the seemingly unrelated regression depend on the mean levels of education of the different family members.

[66] Rouse (1997, Table 3) presents some results which suggest that Ashenfelter and Krueger's findings are attributable to sampling variability associated with their relatively small sample.

Rouse to use one twin's responses about the difference in schooling for the pair as an instrument for the other twin's responses.[67] The IV estimates, presented in the third column of Table 6, are 25% larger than the simple differenced estimates, and about 10% below the corresponding OLS estimates. Rouse (1997) extends the analysis in Ashenfelter and Rouse with one further year of data from the PTS. Her findings, summarized in row 2 of Table 6, are generally consistent with those in Ashenfelter and Rouse (1998), although Rouse's IV estimate is somewhat above the estimate reported by Ashenfelter and Rouse, and actually exceeds the OLS estimate for the same sample.[68]

The study by Miller et al. (1995) uses data for 1170 Australian twin pairs (about one-half female). The advantage of the large sample size is offset by the absence of useable income data: Miller et al. have to impute incomes based on two-digit occupation. Thus, twins with the same two-digit occupation are coded as having the same income.[69] For identical twins Miller et al. (1995) find that the within-family estimate of the return to education is almost 50% lower than the cross-sectional estimate; for fraternal twins, the within-family estimator is 40% lower. Like the PTS, the Australian twins dataset includes multiple reports of each twin's education. Miller et al. (1995) follow Ashenfelter and Krueger's (1994) procedure of using one twin's responses on the difference in schooling for the pair as an instrument for the other's responses. For identical twins, the resulting IV estimate is about 40% above the differenced OLS estimate, but still 25% below the cross-sectional estimate. For fraternal twins the IV estimate is actually slightly above the OLS estimate.

Behrman et al. (1994) analyze a dataset that pools the NAS-NRC sample of white male World War II veterans with data on men from the Minnesota Twins Registry.[70] While the main focus of their paper is on models of inter-familial resource allocation, an appendix table reports cross-sectional and within-family estimates of the return to schooling. For identical twins, Behrman et al. (1994) find that the within-family estimate of the return to schooling is about 50% as large as the cross-sectional OLS estimate,[71] while for fraternal

[67] With two measures of each twin's education there are four possible estimates of the differences in education. Ashenfelter and Krueger (1994) and Rouse (1997) examine the covariance structure of these differences and conclude that the measurement errors in a given twin's reports of her own education and her sibling's education are slightly correlated. Differences in the reports that a given twin provides of the two education levels will eliminate this correlation.

[68] The IV estimate for Rouse (1997) in Table 6 (which uses one twin's report of the difference in the pair's education as an instrument for the other twin's) is not reported in her paper, but was reported by Rouse in a private communication to Gary Solon.

[69] It would be interesting to compare the use of actual income data and imputed incomes in a dataset that includes both, such as the PTS, to judge whether the imputation differentially affects cross-sectional versus within-family estimates of the return to education.

[70] The NAS-NRC sample has been extensively analyzed by some of the same co-authors, e.g., Behrman et al. (1980). Behrman et al. impute earnings for the Minnesota sample using occupation.

[71] This ratio is slightly higher than the ratio reported in earlier work by Behrman et al. (1980) for identical twins in the NAS-NRC sample. Griliches (1979) characterized their results as showing a 65% reduction in the return to schooling between the OLS and within-family estimators.

twins the relative ratio is 80%. Although they do not actually estimate IV models to correct for measurement error, Behrman et al. (1994) report that the reliability of the within-family difference in schooling for identical twins in the NAS-NRC sample is 0.62. Using this estimate, a corrected estimate of the within-family return to schooling for identical twins is 0.056. Behrman et al. (1994) do not give a comparable estimate of the reliability ratio for fraternal twins. Results in Miller et al. (1995) and Ashenfelter and Krueger (1994), however, suggest that the reliability of within-family differences in schooling for fraternal twins is about 0.8. Using this estimate, a corrected estimate of the within-family return to schooling for fraternal twins is 0.071. The relative magnitudes of the OLS and within-family estimators for identical and fraternal twins in Behrman et al. (1994) and Miller et al. (1995) are therefore very comparable.

Finally, Isacsson (1997) analyses earnings and schooling differences among a large sample of Swedish twins (about one-half women). For a subsample of the data he has information on two measures of schooling: one in a register held by Statistics Sweden; another based on self-reported education qualifications.[72] As shown in Table 6, Isacsson finds that the within-family estimate of the return to schooling for identical twins in the subsample with two schooling measures is less than 50% as large as the corresponding OLS estimator, while for fraternal twins the ratio is 80%. He constructs IV estimates for the within-family model using the difference in the survey measures of schooling as an instrument for the differences in the registry measures.[73] For identical twins, the within-family IV estimator is only marginally above the within-family OLS estimate, implying almost no measurement error bias. For fraternal twins, on the other hand, the IV procedure raises the within-family estimate by 35%. Since one would have expected a bigger measurement error attenuation for identical twins than fraternal twins, the patterns of Isacsson's findings are somewhat puzzling.

Isacsson (1997) also constructs measurement-error-corrected estimates of the return to education for a broader sample of twins, assuming "low" and "high" estimates of the reliability of his main schooling measure (reliabilities of 0.85 and 0.95, respectively). The results are summarized in the last two rows of Table 6. For fraternal twins the corrected within-family estimates lie in a fairly tight range (0.044–0.060) that brackets the within-family IV estimate based on the two schooling measures (0.054). For identical twins the range of the corrected estimates is wider (0.027–0.060) and lies above the within-family IV estimate based on the two schooling measures (0.024).

Taken as a whole, Isacsson's results suggest that the measurement-error-corrected within-family estimate of the return to education for *fraternal* twins in Sweden is about as big or even bigger than the corresponding OLS estimate. The precise relative magnitude of the measurement-error-corrected within-family estimate for *identical* twins is more

[72] There is a substantial difference in timing in the two measures. The register-based estimate pertains to 1990 while the self-reported measures were collected in 1974. Isacsson's earnings data are based on administrative records for 1987, 1990, and 1993.

[73] He also reports some evidence on the appropriateness of the assumptions that are needed to justify consistency of these estimates.

Table 6
Cross-sectional and within-family differenced estimates of the return to education for twins[a]

Author	Sample and specification		Cross-sectional OLS	Differenced	
				OLS	IV
1. Ashenfelter and Rouse (1998)	1991–1993 Princeton Twins Survey. Identical male and female twins. Controls Basic controls include quadratic in age, gender and race. Added controls include tenure, marital status and union status.	Basic	0.110 (0.010)	0.070 (0.019)	0.088 (0.025)
		Basic + added controls	0.113 (0.010)	0.078 (0.018)	0.100 (0.023)
2. Rouse (1997)	1991–1995 Princeton Twins Survey. Identical male and female twins. Basic controls as above.		0.105 (0.008)	0.075 (0.017)	0.110 (0.023)
3. Miller et al. (1995)	Australian Twins Register. Identical and fraternal twins. Controls include quadratic in age, gender, marital status. Incomes imputed from occupation	Identical twins	0.064 (0.002)	0.025 (0.005)	0.048 (0.010)
		Fraternal twins	0.066 (0.002)	0.045 (0.005)	0.074 (0.008)
4. Behrman et al. (1994)	NAS-NRC white male twins born 1917–1927, plus male twins born 1936–1955 from Minnesota Twins Registry. Controls include quadratic in age[b]	Identical twins	0.071 (0.002)	0.035 (0.005)	0.056 –
		Fraternal twins	0.073 (0.003)	0.057 (0.005)	0.071 –
5. Isacsson (1997)	Swedish same-sex twins with both administrative and survey measures of schooling. Controls include sex, marital status, quadratic in age, and residence in a large city[c]	Identical twins	0.049 (0.002)	0.023 (0.004)	0.024 (0.008)
		Fraternal twins	0.051 (0.002)	0.040 (0.003)	0.054 (0.006)

Table 6 (*continued*)

Author	Sample and specification		Cross-sectional OLS	Differenced	
				OLS	IV
6. Isacsson (1997)	Swedish same-sex twins. Controls as above[d]	Identical twins	0.046 (0.001)	0.022 (0.002)	0.027 / 0.060 (0.003) (0.007)
		Fraternal twins	0.047 (0.002)	0.039 (0.002)	0.044 / (0.060 (0.002) (0.003)

[a] Notes: See text for sources and information on individual studies.

[b] These authors do not report IV estimates. However, they report an estimate of the reliability of the difference in schooling for identical twins of 0.62 (Berman et al., 1994, Table 4). The IV estimate for identical twins is the differenced estimate divided by 0.62. They do not report an estimate of reliability for fraternal twins. The IV estimate for fraternal twins is the differenced estimate divided by 0.80.

[c] The difference in registry-based estimates of schooling is instrumented by the difference in survey-based measures.

[d] The IV estimates reported in this row are constructed by assuming that the reliability of schooling is 0.095 (yielding the low estimates of the within-family return to schooling) or 0.085 (yielding the high estimates).

uncertain, and seems to be very sensitive to assumptions about measurement error. A cautious interpretation of Isacsson's findings is that there may be some upward bias in OLS estimates of the return to schooling relative to the within-family estimate for identical twins.

What general conclusions can be drawn from the recent twins literature? Suppose on a priori grounds one believes that identical twins have identical abilities. Then the within-family estimator for identical twins, corrected for measurement error biases, is consistent for the average marginal return to schooling in the overall twins population.[74] Assuming that this is the case, the estimates in Table 6 suggest that a cross-sectional OLS estimator yields a slightly upward-biased estimate of the average marginal return to education: the magnitude of the bias ranges across studies from 50% (Isacsson) to zero (Rouse, 1997). Given the limitations of the imputed earnings data used by Miller et al. (1995) and Behrman et al. (1994), and the uncertainties in the measurement error corrections for Isacsson's study, I put more weight on the Ashenfelter–Rouse and Rouse studies, which suggest a smaller range of biases – more like 10–15%.

A second conclusion emerges from the three studies that present results for fraternal twins. In these studies the measurement-error-corrected within-family estimator of the return to education for fraternal twins is about equal to the corresponding OLS estimator. Interestingly, Ashenfelter and Zimmerman's measurement-error-corrected estimate of the return to schooling for brothers – constructed under the assumption that brothers have identical abilities – is also about equal to the corresponding OLS estimate. Since fraternal twins are essentially brothers (or sisters) with the same age, the similarity of the findings for fraternal twins and brothers is reassuring. Assuming that OLS estimates are upward-biased relative to the true average causal effect of education, the within-family estimates based on fraternal twins or brothers must also be upward-biased. Moreover, since the OLS estimator is downward biased by measurement error, whereas the corrected within-family estimates for fraternal twins or brothers are not, one can conclude that the ability bias in within-family estimators for fraternal twins or brothers are *smaller* than the ability bias in cross-sectional OLS estimators: on the order of one-half as large.[75] This implies that ability differences between brothers or sisters are relatively less important determinants of within-family schooling outcomes than are overall ability differences in the determination of schooling outcomes for the population as a whole.

Such a finding opens up the interesting question of how and why families affect the schooling decisions of children with differential abilities. Behrman et al. (1982) present a model incorporating parental preferences in the distribution of education resources across siblings that is consistent with either reinforcing or compensatory behavior (i.e., families may spend more educating either their more- or less-able children). Their empirical findings support the notion of compensatory parental behavior – behavior that would lead to a

[74] The "twins population" may be fairy broad or very narrow, depending on the dataset.

[75] Write the plim of the OLS estimator as $R_0(\bar{\beta} + G_{ols})$ and the plim of the measurement-error corrected fraternal-twins-based estimator as $(\bar{\beta} + G_f)$. If the OLS and corrected fraternal twins estimators are about equal, then $R_0(\bar{\beta} + G_{ols}) = (\bar{\beta} + G_f)$. Assuming that $R_0 \approx 0.9$, if G_f is 10–20% of $\bar{\beta}$, G_f/G_{ols} is 45–60%.

reduction in the relative importance of ability differences in determining education outcomes within families than between families.[76]

If one does not believe that identical twins have identical abilities, then even the within-family estimator of the return to education for identical twins may be biased by ability differences. Ashenfelter and Rouse (1998) present a variety of indirect evidence in support of the hypothesis that identical twins are truly identical, and that differences in their schooling levels are attributable to random factors rather than to ability differences. For example, they report that schooling differences among identical twins are uncorrelated with birth order and with their spouse's education.[77] Despite this evidence, and the strong intuitive appeal of the "equal abilities" assumption for identical twins, however, I suspect that observers with a strong a priori belief in the importance of ability bias will remain unconvinced.

4.4. Direct evidence on the heterogeneity in returns to education

A final set of results in the recent literature that are worth briefly reviewing concern *observable* sources of variation in the return to education. Among the potential sources of heterogeneity that have been identified and studied are school quality, family background, and ability, as measured by IQ or aptitude test scores.

Much interest in the connection between school quality and the return to education was stimulated by the observation that black men had substantially lower returns to schooling than white men in the early 1960s (e.g., Welch, 1973). Moreover, most of the convergence in black-white relative wages that occurred in the 1960s and 1970s can be attributed to a combination of rising relative returns to education for more recent cohorts of black men, and the increasing relative education of blacks relative to whites (Smith and Welch, 1986; 1989). Since the relative quality of schools attended by black students in the segregated southern states improved significantly between 1920 and 1960 (Card and Krueger, 1992b), these facts have led researchers to speculate that increases in school quality may lead to increasing educational attainment and higher returns to education.

Card and Krueger (1992a,b) estimate rates of return to schooling for different cohorts of white and black men who were born in different states and correlate these returns with measures of school quality by cohort and state-of-birth.[78] A distinctive feature of measured returns to education which complicates this analysis is the fact that education-related wage differentials are higher in some parts of the US than others.[79] Card and Krueger address

[76] The more recent study by Behrman et al. (1994), however, finds reinforcing behavior in the allocation of school resources within families.

[77] The latter finding seems to be at odds with results in Behrman et al. (1994, Table A2) who report a strong relationship between differences in education and differences in spouses' education among identical twins.

[78] Previous studies that model school quality effects on the return to education include Akin and Garfinkel (1980) and Link et al. (1980).

[79] This feature of the US labor market was documented in Chiswick (1974). Dahl (1997) presents a thorough summary of the variation in returns to education by state in 1980 and 1990, and evaluates the contribution of selective migration to these patterns.

this by assuming an additive structure to the return to education: an individual born in one state and working in another receives the sum of a state-of-birth component (that presumably varies with school quality); and a state-of-residence component.[80] Under this assumption, Card and Krueger (1992a,b) show that the state-of-birth components in the returns to schooling are systematically correlated with characteristics of the school system. For example, their results suggest that lowering the state-wide pupil-teacher ratio by 10 students raises the rate of return to education earned by students from the state by about 0.9 percentage points.

From the point of view of the models presented in Section 3, another interesting finding reported by Card and Krueger is that students who grew up in states with better quality schools acquired more education. For example, their results for white men imply that a reduction in the statewide pupil-teacher ratio by 10 students raises average educational attainment by 0.6 years. In principle, school quality may affect educational attainment by lowering the marginal cost of schooling, or by raising the marginal benefits of schooling, or both. If one ignores the cost effect, then the implied estimate of the parameter k in Eq. (4) for white men born in the 1920–1950 period is 0.013.[81] This in turn suggests that the magnitude of the endogeneity component ($\psi_0\bar{S} = kf\bar{S}$) in the OLS estimate of the return to schooling is about $0.15f$, where f is the fraction of the variance in school outcomes that is attributable to differences in ability versus differences in tastes. Assuming that the endogeneity bias is about 0.015 (as implied by the results in Ashenfelter and Rouse, 1998) f is about 10%. These calculations are obviously speculative: nevertheless, they illustrate the potential usefulness of data on observable determinants of the return to education in developing a better understanding about the causal effects of education.

Further evidence on the extent of heterogeneity in returns to education and its relationship to school quality and family background is presented in a series of papers by Altonji and Dunn (1995, 1996a,b) that study earnings and schooling data for sibling pairs in the National Longitudinal Surveys of Young Men and Young Women and the Panel Study of Income Dynamics. Altonji and Dunn fit models of log earnings that include education, various control variables, and interactions of education with parental education, IQ, and school quality characteristics.[82] They estimate these models excluding and including family fixed effects. The latter specifications are perhaps the most interesting aspect of their work, since in these models the direct or main effects of family background are held

[80] This assumption is criticized by Heckman et al. (1996) because it ignores the possibility of selective migration. Interestingly, Heckman et al. find larger average effects of school quality in models that control for selective migration by including a function of the distance that individuals have migrated between their state of residence and state of birth. See Card and Krueger (1996) for a summary and discussion.

[81] In the model, $S_i = (b_i - r_i)/k$. If a rise in school quality that raises the average return to schooling by 0.009 leads to 0.6 years of added schooling then $k \approx 0.013$.

[82] There is an earlier literature that includes interactions of family background and ability measures with schooling. Hauser (1973) found little evidence that father's occupational status affected the return to schooling of sons. Similarly, Olneck (1979) concludes that IQ and father's education have little systematic effect on the return to education. On the other hand, Hause (1972) and Willis and Rosen (1979) find positive interactions between aptitude test scores and education.

constant. As one would expect from the discussion of sibling and twin models in Section 3, measurement error plays a potentially important role in the within-family models: Altonji and Dunn develop estimates of the likely magnitude of the attenuation biases that arise in these models and interpret their estimates accordingly.

Altonji and Dunn's results suggest that higher school quality, as measured by spending per pupil, average teacher salaries, or a composite index, raises the return to education. With respect to family background and ability their results are less conclusive. In some of their models that include family fixed effects they find that higher mother's education raises the return to education, although in other samples and specifications the effects are weak and even opposite-signed. Like the earlier literature, they find small and unsystematic effects of parental education on the returns to education in models that exclude family fixed effects. The effects of IQ on the return to education are generally positive (but imprecisely estimated) in the within-family models but negative in the models that exclude family effects.

Ashenfelter and Rouse (1998) also analyze the effects of family background on the returns to schooling for identical twins. Consistent with Altonji and Dunn, their estimates of the interactions between parental education and the difference in schooling between identical twins are positive but imprecise.

Finally, Ashenfelter and Rouse (1998) present some interesting evidence on the existence of declining marginal returns to schooling (i.e., concavity in the relationship between log earnings and schooling at the individual level). They augment a simple within-family differenced earnings equation for identical twins with an interaction between the twins' average education and their difference in education. In the context of the model represented by Eq. (15) the coefficient on this interaction is an estimate of the coefficient k_1.[83] Ashenfelter and Rouse find that returns to schooling decline with the average level of schooling – from about 0.12 at 9 years of schooling to 0.08 at 16 years of schooling – although the gradient is not precisely measured. Such direct evidence of a declining marginal return to schooling supports the interpretation of the IV estimators in Table 4 as yielding estimates of the marginal return to schooling for people who would otherwise have below-average schooling outcomes (relative to the population analyzed in each study).

This brief review suggests three main conclusions. First, the return to education is related to some observable covariates, such as race, school quality, family background measures, and perhaps measured ability. Second, factors such as race, school quality, and mother's education that are associated with higher *returns* to education are also generally associated with higher *levels* of education. These patterns are compatible with an optimizing model of school quality in which individuals are more likely to choose higher levels of education if the return to education is higher. Third, but more tentatively, individual returns to education are declining with the level of education.

[83] Assuming equal abilities for identical twins, Eq. (15) implies that $\Delta \log y_i = b_i \Delta S_i + 1/2 k_1 (S_{i1}^2 - S_{i2}^2) \approx b_i \Delta S_i - k_1 \bar{S}_i \Delta S_i$, where $\bar{S}_i$ is the average education of the twins in family i.

5. Conclusions

Taken as a whole, I believe that the recent literature on the returns to education points to five key conclusions:

1. Consistent with the summary of the literature from the 1960s and 1970s by Griliches (1977, 1979) the *average* (or average marginal) return to education in a given population is not much below the estimate that emerges from a simple cross-sectional regression of earnings on education. The "best available" evidence from the latest studies of identical twins suggests a small upward bias (on the order of 10%) in the simple OLS estimates.
2. Estimates of the return to schooling based on comparisons of brothers or fraternal twins contain some positive ability bias, but less than the corresponding OLS estimates. Ability differences appear to exert relatively less influence on within-family schooling differences than on between-family differences.
3. IV estimates of the return to education based on family background are systematically higher than corresponding OLS estimates and probably contain a bigger upward ability bias than the OLS estimates.
4. Returns to education vary across the population with such observable factors as school quality and parental education.
5. IV estimates of the return to education based on interventions in the school system tend to be 20% or more above the corresponding OLS estimates. While there are several competing explanations for this finding, one plausible hypothesis is that the marginal returns to schooling for certain subgroups of the population – particularly those subgroups whose schooling decisions are most affected by structural innovations in the schooling system – are somewhat higher than the average marginal returns to education in the population as a whole.

While research over the past decade has made genuine progress on the question of the causal effect of education, it may be useful to conclude with a brief list of related topics that have not been as thoroughly addressed. One unresolved question is whether the private return to education – which is the focus of the microeconometric work surveyed here – is equal to, bigger, or smaller than the social return. This question lay at the center of the growth accounting controversy that stimulated much of the modern literature on the return to education, and has re-emerged in the past decade with the return of interest in sources of long-run economic growth. Indeed, much of the "new" growth theory focusses on the possible existence of significant externalities to education.[84] The study of market-level externalities is obviously more difficult than the study of individual-level private returns to education: there are no "identical twins" at the market level. Nevertheless, some of the ideas that underlie the quasi-experimental studies of the private return to education

[84] See the chapter by Topel in this volume for a summary of the this literature with an emphasis on human capital issues.

may be useful at the more aggregate level. For example, institutional changes in the school system may lead to shifts in the relative supply of better-educated workers in one area relative to another that can be used to construct market-level quasi-experimental contrasts.

A second (and related) question is whether the private return to education operates through a homogeneous shift in the productivity of better-educated workers, or through a more complex mechanism, such as differential access to different types of jobs. Some authors interpret the research on sheepskin effects described in Section 2 as distinguishing between these alternatives (see, e.g., Weiss, 1995). An innovative study of the market returns to a General Educational Development (GED) certificate by Tyler et al. (1998) suggests that credentials per se have a significant value in the US labor market, while other work (e.g., Cameron and Heckman, 1993) has questioned this hypothesis.[85]

A third question that has received renewed interest in the recent literature is how returns to education vary with observable characteristics, such as family background, school quality, ability, or location. One worthy goal of future research is to develop a better understanding of the extent to which the effects of permanent characteristics like family background on the returns to education "explain" their effects on educational attainment. A loftier goal is to understand the joint determination of schooling attainment and other endogenous outcomes like location or occupation in the context of a structural model of schooling and earnings determination.

A final issue that I have ignored in this chapter is variation in the returns to education over time: either for the economy as a whole, or for fixed cohorts of individuals. Over the past 15 years the conventionally measured return to education has risen by 35–50% (see Autor et al., 1997, or the chapter by Katz and Autor in this volume). Relative to these shifts, the ability biases that are the focus of the literature reviewed here seem very modest in magnitude. Nevertheless, some authors have argued that changes over time in the overall return to education may be driven in part by *changes* in the magnitude of the ability bias components (e.g., Taber, 1998; Cawley et al., 1998). Some of the methods developed to study the extent of ability bias in a cross-sectional dataset can be extended to panel data, offering the possibility of modelling time-varying ability biases.

Appendix A

A.1. OLS estimation of a random coefficients model

Let y denote a (scalar) outcome variable that is related to a k-dimensional covariate X

[85] Tyler et al.'s (1998) research design underscores the value of detailed institutional knowledge in helping to untangle causal mechanisms in the labor market. The GED certificate is awarded in lieu of high school graduation for successful completion of a test. In some states, however, the required test score to earn a GED is lower, allowing one to test whether the certificate itself is rewarded in the labor market, or only the underlying "knowledge". Tyler et al.'s results suggest that the certificate itself is important, since people with the same scores who would earn the degree in one state but not another appear to earn more when they have the degree.

through a linear regression model with random intercept α and random slope coefficients β:

$$y = \alpha + X'\beta + u = \bar{\alpha} + X'\bar{\beta} + (\alpha - \bar{\alpha}) + X'(\beta - \bar{\beta}) + u, \tag{A.1}$$

where $\bar{\alpha}$ and $\bar{\beta}$ denote the means of α and β, respectively, and $\mathrm{E}[X'u] = 0$. Denote the linear projections of α and β on X by

$$\alpha - \bar{\alpha} = \lambda'(X - \bar{X}) + v_1, \tag{A.2a}$$

$$\beta - \bar{\beta} = \psi(X - \bar{X}) + v_2, \tag{A.2b}$$

where $\mathrm{E}[X'v_1] = \mathrm{E}[X'v_2] = 0$ (by definition of λ and ψ). Using these definitions,

$$\mathrm{E}[(\beta - \bar{\beta})(X - X')] = \psi\mathrm{E}[(X - \bar{X})(X - \bar{X})'] = \psi\mathrm{var}[X],$$

and therefore

$$\mathrm{cov}[X, X'(\beta - \bar{\beta})] = \mathbf{E}[(X - \bar{X})(\beta - \bar{\beta})'X] = \mathrm{E}[(X - \bar{X})(\beta - \bar{\beta})'(\bar{X} + (X - \bar{X}))]$$

$$= \mathrm{var}[X]\psi'\bar{X} + D,$$

where

$$D = \mathrm{E}[(X - \bar{X})(X - \bar{X})'(\beta - \bar{\beta})].$$

The probability limit of the OLS estimator of $\bar{\beta}$ for Eq. (A.1) is therefore

$$\mathrm{var}[X]^{-1}\mathrm{cov}[X, y] = \mathrm{var}[X]^{-1}\{\mathrm{var}[X]\beta + \mathrm{var}[X]\lambda + \mathrm{var}[X]\psi'\bar{X} + D\}$$

$$= \bar{\beta} + \lambda + \psi'\bar{X} + \mathrm{var}[X]^{-1}D.$$

Notice that if X and β are jointly symmetrically distributed then $D = 0$. In this case the probability limit of the OLS regression coefficient is just

$$\mathrm{plim}(\beta_{\mathrm{ols}}) = \bar{\beta} + \lambda + \psi'\bar{X}. \tag{A.3}$$

A.2. Estimation of a random coefficients model

Consider the estimation of Eq. (A.1) when a set of instruments Z is available with the property that

$$\mathrm{E}[(\alpha - \bar{\alpha}) \mid Z] = 0, \tag{A.3a}$$

$$\mathrm{E}[(\beta - \bar{\beta}) \mid Z] = 0. \tag{A.3b}$$

Assuming that Z includes a vector of constants, denote the reduced form projection of X on Z by

$$X = \Pi Z + v. \tag{A.4}$$

Finally, assume that

$$E[u \mid X, Z] = 0, \qquad E[v \mid Z] = 0, \tag{A.5a}$$

$$E[\alpha - \bar{\alpha} \mid X, Z] = \lambda_x' X + \lambda_z' Z, \tag{A.5b}$$

$$E[\beta - \bar{\beta} \mid X, Z] = \psi_x X + \psi_z Z. \tag{A.5c}$$

Assumption (A.5a) strengthens the orthogonality conditions defining the error components u and v into assumptions on conditional expectations. Assumptions (A.5b) and (A.5c) specify that the conditional expectations of α and β are linear in X and Z. Under these assumptions,

$$0 = E[\alpha - \bar{\alpha} \mid Z] = E[\lambda_x'(\Pi Z + v) + \lambda_z' Z \mid Z] = (\lambda_x'\Pi + \lambda_z')Z,$$

implying that $\lambda_z' = -\lambda_x'\Pi$. Similarly

$$0 = E[\beta - \bar{\beta} \mid Z] = E[\psi_x(\Pi Z + v) + \psi_z Z \mid Z] = (\psi_x\Pi + \psi_z)Z,$$

implying that $\psi_z = -\psi_x\Pi$. Substituting (A.5b) and (A.5c) into (A.1) and taking expections conditional on (X,Z) yields

$$\begin{aligned} E[y \mid X, Z] &= \bar{\alpha} + X'\bar{\beta} + \lambda_x' X + \lambda_z' Z + X'(\psi_x X + \psi_z Z) \\ &= \bar{\alpha} + X'\bar{\beta} + \lambda_x(X - \Pi Z) + X'\psi_x(X - \Pi Z) \\ &= \bar{\alpha} + X'\bar{\beta} + \lambda_x v + X'\psi_x v. \end{aligned} \tag{A.6}$$

Using standard arguments, (A.6) implies that consistent estimates of $\bar{\beta}$ can be obtained from a "control function" estimator that includes X, $\check{v}$ (the residual from a regression of X on Z) and interactions of X and $\check{v}$ (see also Garen, 1984). Notice that if β is constant then the control function is simply $\check{v}$, yielding the conventional IV estimator. (In this case the preceding assumptions can be weakened by replacing the expectations operator in Eqs. (A.3) and (A.5) by the linear projection operator.)

A.3. Measurement error in a bivariate regression model

Consider a bivariate regression model

$$y = X_1 b_1 + X_2 b_2 + u, \tag{A.7}$$

where X_1 and X_2 are measured with error. Denote the observed value of X_i by X_i^o $(i = 1,2)$, and assume that

$$X_1^o = X_1 + \varepsilon_1, \qquad X_2^o = X_2 + \varepsilon_2,$$

where $E[X_i \varepsilon_j] = E[\varepsilon_1 \varepsilon_2] = 0$. Let R_1 and R_2 denote the reliability ratios of X_1 and X_2, respectively, where

$R_i \equiv \text{cov}[X_i^o, X_i]/\text{var}[X_i^o]$.

Finally, consider the auxiliary regressions

$$X_1 = X_1^o a_{11} + X_2^o a_{12} + v_1, \tag{A.8a}$$

$$X_2 = X_1^o a_{21} + X_2^o a_{22} + v_2, \tag{A.8b}$$

where v_i is orthogonal to X_1^o and X_2^o for $i = 1,2$. The coefficients of these regressions can be expressed in terms of the variances of the observed X's, the reliability ratios, and ρ, the correlation of the observed covariates, X_1^o and X_2^o. If y is regressed on the observed X's:

$$y = X_1^o c_1 + X_2^o c_2 + e,$$

the regression coefficients will equal

$$c_1 = b_1 a_{11} + b_2 a_{21}, \qquad c_2 = b_1 a_{12} + b_2 a_{22}.$$

It is easy to show that

$$c_1 = b_1 \frac{R_1 - \rho^2}{1 - \rho^2} + b_2 \frac{1 - R_2}{1 - \rho^2} \times \frac{\text{cov}[X_1^o, X_2^o]}{\text{var}[X_1^o]}, \tag{A.9a}$$

$$c_2 = b_2 \frac{R_2 - \rho^2}{1 - \rho^2} + b_1 \frac{1 - R_1}{1 - \rho^2} \times \frac{\text{cov}[X_1^o, X_2^o]}{\text{var}[X_2^o]}, \tag{A.9b}$$

References

Akin, John S. and Irwin Garfinkel (1980), "The quality of education and cohort variation in black-white earnings differentials: comment", American Economic Review 70: 186–191.

Altonji, Joseph G. and Thomas A. Dunn (1995), "The effects of school and family characteristics on the return to education", Working paper no. 5072 (NBER, Cambridge, MA).

Altonji, Joseph G. and Thomas A. Dunn (1996a), "Using siblings to estimate the effect of school quality on wages", Review of Economics and Statistics 78: 665–671.

Altonji, Joseph G. and Thomas A. Dunn (1996b), "The effects of family characteristics on the return to schooling", Review of Economics and Statistics 78: 692–704.

Angrist, Joshua D. and Guido W. Imbens (1995), "Two-stage least squares estimation of average causal effects in models with variable treatment intensity", Journal of the American Statistical Association 90: 431–442.

Angrist, Joshua D. and Alan B. Krueger (1991), "Does compulsory school attendance affect schooling and earnings", Quarterly Journal of Economics 106: 979–1014.

Angrist, Joshua D. and Alan B. Krueger (1992), "Estimating the payoff to schooling using the Vietnam-era draft lottery", Working paper no. 4067 (NBER, Cambridge, MA).

Angrist, Joshua D. and Alan B. Krueger (1995), "Split sample instrumental variables estimates of the return to schooling", Journal of Business and Economics Statistics 13: 225–235.

Angrist, Joshua D. and Victor Lavy (1997), "Using Maimonides' rule to estimate the effect of class size on scholastic achievement", Working paper no. 5888 (NBER, Cambridge, MA).

Angrist, Joshua D. and Whitney K. Newey (1991), "Over-identification tests in earnings functions with fixed effects", Journal of Business and Economic Statistics 9: 317–323.

Angrist, Joshua D., Guido W. Imbens and Donald B. Rubin (1996), "Identification of causal effects using instrumental variables", Journal of the American Statistical Association 91: 444–472.

Ashenfelter, Orley and Colm Harmon (1998), "Editors introduction", Labour Economics (special issue on education), in press.

Ashenfelter, Orley and Alan B. Krueger (1994), "Estimates of the economic return to schooling for a new sample of twins", American Economic Review 84: 1157–1173.

Ashenfelter, Orley and Cecilia E. Rouse (1998), "Income, schooling and ability: evidence from a new sample of identical twins", Quarterly Journal of Economics 113: 253–284.

Ashenfelter, Orley and David Zimmerman (1997), "Estimates of the return to schooling from sibling data: fathers, sons and brothers", Review of Economics and Statistics 79: 1–9.

Autor, David H., Lawrence F. Katz and Alan B. Krueger (1997), "Computing inequality: have computers changed the labor market?" Working paper no. 5956 (NBER, Cambridge, MA).

Becker, Gary S. (1964), Human capital: a theoretical and empirical analysis, with special reference to education (Columbia University Press, New York).

Becker, Gary S. (1967), Human capital and the personal distribution of income (University of Michigan Press, Ann Arbor, MI).

Behrman, Jere R., Z. Hrubec, Paul Taubman and Terence J. Wales (1980), Socioeconomic success: a study of the effects of genetic endowments, family environment and schooling (North Holland, Amsterdam).

Behrman, Jere R., Robert Pollack and Paul Taubman (1982), "Parental preferences and the provision of progeny", Journal of Political Economy 90: 52–73.

Behrman, Jere R., Mark R. Rosenzweig and Paul Taubman (1994), "Endowments and the allocation of schooling in the family and the marriage market: the twins experiment", Journal of Political Economy 102: 1131–1174.

Belman, Dale and John Heywood (1991), "Sheepskin effects in the return to education", Review of Economics and Statistics 73: 720–724.

Bound, John and David A. Jaeger (1996), "On the validity of season of birth as an instrument in wage equations: a comment on Angrist and Krueger's 'Does compulsory school attendance affect schooling and earnings?'", Working paper no. 5835 (NBER, Cambridge, MA).

Bound, John, David A. Jaeger and Regina M. Baker (1995), "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak", Journal of the American Statistical Association 90: 443–450.

Butcher, Kristin F. and Anne Case (1994), "The effect of sibling composition on women's education and earnings", Quarterly Journal of Economics 109: 531–563.

Cameron, Stephen V. and James J. Heckman (1993), "The non-equivalence of high school equivalents", Journal of Labor Economics 11: 1–47.

Card, David (1995a), "Earnings, schooling and ability revisited", in: Solomon Polachek, ed., Research in labor economics, Vol. 14 (JAI Press, Greenwich, CT) pp. 23–48.

Card, David (1995b), "Using geographic variation in college proximity to estimate the return to schooling", in: Louis N. Christofides, E. Kenneth Grant and Robert Swidinsky, eds., Aspects of labour market behaviour: essays in honour of John Vanderkamp (University of Toronto Press, Toronto, Canada) pp. 201–222.

Card, David and Alan B. Krueger (1992a), "Does school quality matter: returns to education and the characteristics of public schools in the United States", Journal of Political Economy 100: 1–40.

Card, David and Alan B. Krueger (1992b), "School quality and black-white relative earnings: a direct assessment", Quarterly Journal of Economics 107: 151–200.

Card, David and Alan B. Krueger (1996), "Labor market effects of school quality: theory and evidence", in: Gary Burtless, ed., Does money matter? The effect of school resources on student achievement and adult success (Brookings Institution, Washington, DC).

Cawley, John, James J. Heckman and Edward Vytlacil (1998), "Cognitive ability and the rising return to education", Working paper no. 6388 (NBER, Cambridge, MA).

Chamberlain, Gary (1977), "Omitted variable bias in panel data: estimating the returns to schooling", Annals de l'Insee 30–31: 49–82.

Chamberlain, Gary (1982), "Multivariate regression models for panel data", Journal of Econometrics 18: 5–46.
Chamberlain, Gary and Zvi Griliches (1975), "Unobservables with a variance components structure: ability, schooling and the economic success of brothers", International Economic Review 16: 422–449.
Chamberlain, Gary and Zvi Griliches (1977), "More on brothers", in: Paul Taubman, ed., Kinometrics (North Holland, Amsterdam).
Chiswick, Barry (1974), Income inequality: regional analysis within a human capital framework (Columbia University Press, New York).
Cohn, Elchanan and John T. Addison (1997), "The economic returns to lifelong learning", Working paper B-97-04 (Division of Research, University of South Carolina College of Business Administration).
Conneely, Karen and Roope Uusitalo (1997), "Estimating heterogeneous treatment effects in the Becker schooling model", Unpublished discussion paper (Industrial Relations Section, Princeton University).
Dahl, Gordon B. (1997), "Mobility and the returns to education: testing a Roy model with multiple markets", Unpublished discussion paper (Industrial Relations Section, Princeton University).
Denison, Edward F. (1964), "Measuring the contribution of education", in: The residual factor and economic growth (OECD, Paris).
Freeman, Richard B. (1986), "Demand for education", in Orley Ashenfelter and Richard Layard, eds., Handbook of labor economics (North Holland, Amsterdam).
Garen, John (1984), "The returns to schooling: a selectivity bias approach with a continuous choice variable", Econometrica 52: 1199–1218.
Goodman, Jerry (1979), "The economic returns of education: an assessment of alternative models", Social Science Quarterly 60: 269–283.
Griliches, Zvi (1970), "Notes on the role of education in production functions and growth accounting", in: W. Lee Hansen, ed., Studies in income and wealth, Vol. 35 (Columbia University Press, New York).
Griliches, Zvi (1977), "Estimating the returns to schooling: some econometric problems", Econometrica 45: 1–22.
Griliches, Zvi (1979), "Sibling models and data in economics: beginnings of a survey", Journal of Political Economy 87: S37–S65.
Harmon, Colm and Ian Walker (1995), "Estimates of the economic return to schooling for the United Kingdom", American Economic Review 85: 1278–1286.
Hause, John C. (1972), "Earnings profile: ability and schooling", Journal of Political Economy 80: S108–S138.
Hauser, Robert M. (1973), "Socioeconomic background and differential returns to education", in: Lewis C. Solomon and Paul J. Taubman, eds., Does college matter? Some evidence on the impacts of higher education (Academic Press, New York).
Hausman, Jerry A. and William E. Taylor (1981), "Panel data and unobservable individual effects", Econometrica 49: 1377–1398.
Heckman, James J. and Solomon Polachek (1974), "Empirical evidence on the functional form of the earnings-schooling relationship", Journal of the American Statistical Association 69: 350–354.
Heckman, James J. and Edward Vytlacil (1998), "Instrumental variables methods for the correlated random coefficient model: estimating the rate of return to schooling when the return is correlated with schooling", Unpublished discussion paper (University of Chicago).
Heckman, James J., Anne Layne-Farrar and Petra Todd (1996), "Human capital pricing equations with an application to estimating the effect of school quality on earnings", Review of Economics and Statistics 78: 562–610.
Hungerford, Thomas and Gary Solon (1987), "Sheepskin effects in the return to education", Review of Economics and Statistics 69: 175–177.
Ichino, Andrea and Rudolf Winter-Ebmer (1998), "The long-run educational cost of World War II", Unpublished discussion paper (European University Institute).
Isacsson, Gunnar (1997), "Estimates of the return to schooling in Sweden from a large sample of twins", Unpublished discussion paper (Center for Research on Transportation and Society, Borlange, Sweden).
Jakubson, George (1988), "The sensitivity of labor supply parameter estimates to unobserved individual effects:

fixed and random effects estimates in a nonlinear model using panel data", Journal of Labor Economics 6: 302–329.
Kane, Thomas J. and Cecilia E. Rouse (1993), "Labor market returns to two- and four-year colleges: is a credit a credit and do degrees matter?" Working paper no. 4268 (NBER, Cambridge, MA).
Kane, Thomas J., Cecilia E. Rouse and Douglas Staiger (1997), "Estimating returns to education when schooling is misreported", Unpublished discussion paper (Industrial Relations Section, Princeton University).
Kessler, Daniel (1991), "Birth order, family size and achievement: family structure and wage determination", Journal of Labor Economics 9: 413–426.
Kominski, Robert and Paul M. Siegel (1992), "Measuring educational attainment in the current population survey" (United States Department of Commerce Bureau of the Census, Population Division, Washington, DC).
Lang, Kevin (1993), "Ability bias, discount rate bias and the return to education", Unpublished discussion paper (Department of Economics, Boston University).
Link, Charles, Edward Ratledge and Kenneth Lewis (1980), "The quality of education and cohort variation in black-white earnings differentials: reply", American Economic Review 70: 196–203.
Maluccio, John (1997), "Endogeneity of schooling in the wage function", Unpublished manuscript (Department of Economics, Yale University).
Miller, Herman P. (1955), Income of the American people (United States Government Printing Office, Washington, DC).
Miller, Herman P. (1966), Income distribution in the United States (United States Government Printing Office, Washington, DC).
Miller, Paul, Charles Mulvey and Nick Martin (1995), "What do twins studies reveal about the economic returns to education? A comparison of Australian and U.S. findings", American Economic Review 85: 586–599.
Mincer, Jacob (1974), Schooling, experience and earnings (Columbia University Press, New York).
Murphy, Kevin M. and Finis Welch (1990), "Empirical age-earnings profiles", Journal of Labor Economics 8: 202–229.
Olneck, Michael R. (1979), "The effects of education", in: Christopher Jencks et al., eds., Who gets ahead? (Basic Books, New York).
Park, Jin Heum (1994), "Returns to schooling: a peculiar deviation from linearity", Working paper no. 335 (Industrial Relations Section, Princeton University).
Park, Jin Heum (1996), "Measuring education over time", Economics Letters 60: 425–428.
Psacharopoulos, George (1985), "Returns to education: a further international update and implications", Journal of Human Resources 20: 583–604.
Psacharopoulos, George (1994), "Returns to investment in education: a global update", World Development 22: 1325–1343.
Rouse, Cecilia E. (1997), "Further estimates of the economic return to schooling from a new sample of twins", Unpublished discussion paper (Industrial Relations Section, Princeton University).
Siebert, W. Stanley (1985), "Developments in the economics of human capital", in: Labour economics (Longman, London).
Siegel, Paul M. and Robert Hodge (1968), "A causal approach to the study of measurement error", in: Hubert Blalock and Ann Blalock, eds., Methodology in social research (McGraw Hill, New York).
Smith, James P. and Finis Welch (1986), Closing the gap: forty years of economic progress for blacks (RAND, Santa Monica, CA).
Smith, James P. and Finis Welch (1989), "Black economic progress after Myrdal", Journal of Economic Literature 27: 519–564.
Staiger, Douglas and James H. Stock (1997), "Instrumental variables regression with weak instruments", Econometrica 65: 557–586.
Taber, Christopher (1998), "The college premium in the eighties: return to college or return to ability", Unpublished discussion paper (Department of Economics, Northwestern University).
Tyler, John H., Richard J. Murnane and John B. Willett (1998), "Estimating the impact of the GED on the

earnings of young dropouts using a series of natural experiments", Working paper no. 6391 (NBER, Cambridge, MA).

Weiss, Andrew (1995), "Human capital vs. signalling explanations of wages", Journal of Economic Perspectives 9: 133–154.

Welch, Finis (1973), "Black-white differences in returns to schooling", American Economic Review 63: 893–907.

Willis, Robert J. (1986), "Wage determinants: a survey and reinterpretation of human capital earnings functions", in: Orley Ashenfelter and Richard Layard, eds., Handbook of labor economics (North Holland, Amsterdam).

Willis, Robert J. and Sherwin Rosen (1979), "Education and self-selection", Journal of Political Economy 79: S7–S36.

Wooldridge, Jeffrey M. (1997), "On two-stage least squares estimation of the average treatment effect in a random coefficient model", Economics Letters 56: 129–133.

Zheng, John (1996), "Specification testing and nonparametric estimation of the human capital model", Unpublished discussion paper (Department of Economics, University of Texas at Austin).

Chapter 31

THE ECONOMICS AND ECONOMETRICS OF ACTIVE LABOR MARKET PROGRAMS

JAMES J. HECKMAN*

University of Chicago

ROBERT J. LALONDE*

University of Chicago

JEFFREY A. SMITH*

University of Western Ontario

Contents

Abstracts 1866
JEL codes 1867
1 Introduction 1867
2 Public job training and active labor market policies 1871
3 The evaluation problem and the parameters of interest in evaluating social programs 1877
3.1 The evaluation problem 1877
3.2 The counterfactuals of interest 1879
3.3 The counterfactuals most commonly estimated in the literature 1882
3.4 Is treatment on the treated an interesting economic parameter? 1886
4 Prototypical solutions to the evaluation problem 1891
4.1 The before–after estimator 1891
4.2 The difference-in-differences estimator 1894
4.3 The cross-section estimator 1896
5 Social experiments 1899
5.1 How social experiments solve the evaluation problem 1899
5.2 Intention to treat and substitution bias 1903
5.3 Social experiments in practice 1905
6 Econometric models of outcomes and program participation 1914
6.1 Uses of economic models 1914
6.2 Prototypical models of earnings and program participation 1914
6.3 Expected present value of earnings maximization 1915
6.4 The role of program eligibility rules in determining participation 1932

* We thank Susanne Ackum Agell for her helpful comments on Scandanavian active labor market programs and Costas Meghir for very helpful comments on Sections 1–7 and Wilbert van der Klaauw for comments on Section 7.4.6.

Handbook of Labor Economics, Volume 3, Edited by A. Ashenfelter and D. Card

6.5 Administrative discretion and the efficiency and equity of training provision 1933
6.6 The conflict between the economic approach to program evaluation and the modern approach to social experiments 1935
7 Non-experimental evaluations 1936
7.1 The problem of causal inference in non-experimental evaluations 1936
7.2 Constructing a comparison group 1938
7.3 Econometric evaluation estimators 1941
7.4 Identification assumptions for cross-section estimators 1950
7.5 Using aggregate time series data on cohorts of participants to evaluate programs 1972
7.6 Panel data estimators 1973
7.7 Robustness to biased sampling plans 1985
7.8 Bounding and sensitivity analysis 1989
8 Econometric practice 1992
8.1 Data sources 1993
8.2 Characterizing selection bias 1998
8.3 A simulation study of the sensitivity of non-experimental methods 2007
8.4 Specification testing and the fallacy of alignment 2025
9 Indirect effects, displacement and general equilibrium treatment effects 2033
9.1 Review of the traditional approaches to displacement and substitution 2035
9.2 General equilibrium approaches 2036
9.3 Summary of general equilibrium approaches 2043
10 A survey of empirical findings 2043
10.1 The objectives of program evaluations 2043
10.2 The impact of government programs on labor market outcomes 2050
10.3 The findings from US social experiments 2054
10.4 The findings from non-experimental evaluations of US programs 2064
10.5 The findings from European evaluations 2069
11 Conclusions 2080
References 2085

Abstract

Policy makers view public sector-sponsored employment and training programs and other active labor market policies as tools for integrating the unemployed and economically disadvantaged into the work force. Few public sector programs have received such intensive scrutiny, and been subjected to so many different evaluation strategies. This chapter examines the impacts of active labor market policies, such as job training, job search assistance, and job subsidies, and the methods used to evaluate their effectiveness. Previous evaluations of policies in OECD countries indicate that these programs usually have at best a modest impact on participants' labor market prospects. But at the same time, they also indicate that there is considerable heterogeneity in the impact of these programs. For some groups, a compelling case can be made that these policies generate high rates of return, while for other groups these policies have had no impact and may have been harmful. Our discussion of the methods used to evaluate these policies has more general interest. We believe that the same issues arise generally in the social sciences and are no easier to address elsewhere. As a result, a major focus of this chapter is on the methodological lessons learned from evaluating these programs. One of the most important of these lessons is that there is no inherent method of choice for conducting program evaluations. The choice between experimental and non-experimental methods

or among alternative econometric estimators should be guided by the underlying economic models, the available data, and the questions being addressed. Too much emphasis has been placed on formulating alternative econometric methods for correcting for selection bias and too little given to the quality of the underlying data. Although it is expensive, obtaining better data is the only way to solve the evaluation problem in a convincing way. However, better data are not synonymous with social experiments.

JEL codes: J24; J31; C50; C93; J64

1. Introduction

Public provision of job training, of wage subsidies and of job search assistance is a feature of the modern welfare state. These activities are cornerstones of European "active labor market policies", and have been a feature of US social welfare policy for more than three decades. Such policies also have been advocated as a way to soften the shocks administered to the labor markets of former East Block and Latin economies currently in transition to market-based systems.

A central characteristic of the modern welfare state is a demand for "objective" knowledge about the effects of various government tax and transfer programs. Different parties benefit and lose from such programs. Assessments of these benefits and losses often play critical roles in policy decision-making. Recently, interest in evaluation has been elevated as many economies with modern welfare states have floundered, and as the costs of running welfare states have escalated.

This chapter examines the evidence on the effectiveness of welfare state active labor market policies such as training, job search and job subsidy policies, and the methods used to obtain the evidence on their effectiveness. Our methodological discussion of alternative approaches to evaluating programs has more general interest. Few US government programs have received such intensive scrutiny, and been subject to so many different types of evaluation methodologies, as has governmentally-supplied job training. In part, this is due to the fact that short-run measures of government training programs are more easily obtained and are more readily accepted. Outcomes such as earnings, employment, and educational and occupational attainment are all more easily measured than the outcomes of health and public school education programs. In addition, short-run measures of the outcomes of training programs are more closely linked to the "treatment" of training. In public school and health programs, a variety of inputs over the lifecycle often give rise to measured outcomes. For these programs, attribution of specific effects to specific causes is more problematic.

A major focus of this chapter is on the general lessons learned from over 30 years of experience in evaluating government training programs. Most of our lessons come from American studies because the US government has been much more active in promoting evaluations than have other governments, and the results from the evaluations are often used to expand – or contract – government programs. We demonstrate that recent studies

in Europe indicate that the basic patterns and lessons from the American case apply more generally.

The two relevant empirical questions in this literature are (i) adjusting for their lower skills and abilities, do participants in government employment and training programs benefit from these programs? and (ii) are these programs worthwhile social investments? As currently constituted, these programs are often ineffective on both counts. For most groups of participants, the benefits are modest, and at worst participation in government programs is harmful. Moreover, many programs and initiatives cannot pass a cost-benefit test. Even when programs are cost effective, they are rarely associated with a large-scale improvement in skills. But, at the same time, there is substantial heterogeneity in the impacts of these programs. For some groups these programs appear to generate significant benefits both to the participants and to society.

We believe that there are two reasons why the private and social gains from these programs are generally small. First, the per-capita expenditures on participants are usually small relative to the deficits that these programs are being asked to address. In order for such interventions to generate large gains they would have to be associated with very large internal rates of return. Moreover, these returns would have to be larger than what is estimated for private sector training (Mincer, 1993). Another reason that the gains from these programs are generally low is that these services are targeted toward relatively unskilled and less able individuals. Evidence on the complementarity between the returns to training and skill in the private sector suggests that the returns to training in the public sector should be relatively low.

We also survey the main methodological lessons learned from thirty years of evaluation activity conducted mainly in the United States. We have identified eight lessons from the evaluation literature that we believe should guide practice in the future. First, there are many parameters of interest in evaluating any program. This multiplicity of parameters results in part because of the heterogeneous impacts of these programs. As a result of this heterogeneity, some popular estimators that are well-suited for estimating one set of parameters are poorly suited for estimating others. The understanding that responses to the same measured treatment are heterogenous across people, that measured treatments themselves are heterogeneous, that in many cases people participate in programs based in part on this heterogeneity and that econometric estimators should allow for this possibility, is an important insight of the modern literature that challenges traditional approaches to program evaluation. Because of this heterogeneity, many different parameters are required to answer the interesting evaluation questions.

Second, there is inherently no method of choice for conducting program evaluations. The choice of an appropriate estimator should be guided by the economics underlying the problem, the data that are available or that can be acquired, and the evaluation question being addressed.

A third lesson from the evaluation literature is that better data help a lot. The data available to most analysts have been exceedingly crude. Too much has been asked of econometric methods to remedy the defects of the underlying data. When certain features

of the data are improved, the evaluation problem becomes much easier. The best solution to the evaluation problem lies in improving the quality of the data on which evaluations are conducted and not in the development of formal econometric methods to circumvent inadequate data.

Fourth, it is important to compare comparable people. Many non-experimental evaluations identify the parameter of interest by comparing observationally different persons using extrapolations based on inappropriate functional forms imposed to make incomparable people comparable. A major advantage of non-parametric methods for solving the problem of selection bias is that, rigorously applied, they force analysts to compare only comparable people.

Fifth, evidence that different non-experimental estimators produce different estimates of the same parameter does not indicate that non-experimental methods cannot address the underlying self-selection problem in the data. Instead, different estimates obtained from different estimators simply indicate that different estimators address the selection problem in different ways and that non-random participation in social programs is an important problem. Different methods produce the same estimates only if there is no problem of selection bias.

Sixth, a corollary lesson, derived from lessons three, four and five, is that the message from LaLonde's (1986) influential study of non-experimental estimators has been misunderstood. Once analysts define bias clearly, compare comparable people, know a little about the unemployment histories of trainees and comparison group members, administer them the same questionnaire and place them in the same local labor market, much of the bias in using non-experimental methods is attenuated. Variability in estimates across estimators arises from the fact that different non-experimental estimators solve the selection problem under different assumptions, and these assumptions are often incompatible with each other. Only if there is no selection bias would all evaluation estimators identify the same parameter.

Seventh, three decades of experience with social experimentation have enhanced our understanding of the benefits and limitations of this approach to program evaluation. Like all evaluation methods, this method is based on implicit identifying assumptions. Experimental methods estimate the effect of the program compared to no programs at all when they are used to evaluate the effect of a program for which there are few good substitutes. They are less effective when evaluating ongoing programs in part because they appear to disrupt established bureaucratic procedures. The threat of disruption leads local bureaucrats to oppose their adoption. To the extent that programs are disrupted, the program evaluated by the method is not the ongoing program that one seeks to evaluate. The parameter estimated in experimental evaluations is often not likely to be of primary interest to policy makers and researchers, and under any event has to be more carefully interpreted than is commonly done in most public policy discussions. However, if there is no disruption, and the other problems that plague experiments are absent, the evidence from social experiments provides a benchmark for learning about the performance of alternative non-experimental methods.

Eighth, and finally, programs implemented at a national or regional level affect both participants and non-participants. The current practice in the entire "treatment effect" literature is to ignore the indirect effects of programs on non-participants by assuming they are negligible. This practice can produce substantially misleading estimates of program impacts if indirect effects are substantial. To account for the impacts of programs on both participants and non-participants, general equilibrium frameworks are required when programs substantially impact the economy.

The remainder of the chapter is organized as follows. In Section 2, we distinguish among several types of active labor market policies and describe the types of employment and training services offered both in the US and in Europe, their approximate costs, and their intended effects. We introduce the evaluation problem in Section 3. We discuss the importance of heterogeneity in the response to treatment for defining counterfactuals of interest. We consider what economic questions the most widely used counterfactuals answer. In Section 4, we present three prototypical solutions to the evaluation problem cast in terms of mean impacts. These prototypes are generalized throughout the rest of this chapter, but the three basic principles introduced in this section underlie all approaches to program evaluation when the parameters of interest are means or conditional means. In Section 5, we present conditions under which social experiments solve the evaluation problem and assess the effectiveness of social experiments as a tool for evaluating employment and training programs. In Section 6, we outline two prototypical models of program participation and outcomes that represent the earliest and the latest thinking in the literature. We demonstrate the implications of these decision rules for the choice of an econometric evaluation estimator. We discuss the empirical evidence on the determinants of participation in government training programs.

The econometric models used to evaluate the impact of training programs in non-experimental settings are described in Section 7. The interplay between the economics of program participation and the choice of an appropriate evaluation estimator is stressed. In Section 8, we discuss some of the lessons learned from implementing various approaches to evaluation. Included in this section are the results of a simulation analysis based on the empirical model of Ashenfelter and Card (1985), where we demonstrate the sensitivity of the performance of alternative estimators to assumptions about heterogeneity in impacts among persons and to other data generating processes of the underlying econometric model. We also reexamine LaLonde's (1986) evidence on the performance of non-experimental estimators and reinterpret the main lessons from his study.

Section 9 discusses the problems that arise in using microeconomic methods to evaluate programs with macroeconomic consequences. A striking example of the problems that can arise from this practice is provided. Two empirically operational general equilibrium frameworks are presented, and the lessons from applying them in practice are summarized. Section 10 surveys the findings from the non-experimental literature, and contrasts them with those from experimental evaluations. We conclude in Section 11 by surveying the main methodological lessons learned from the program evaluation literature on job training.

2. Public job training and active labor market policies

Many government policies affect employment and wages. The "active labor market" policies we analyze have two important features that distinguish them from general policies, such as income taxes, that also affect the labor market. First, they are targeted toward the unemployed or toward those with low skills or little work experience who have completed (usually at a low level) their formal schooling. Second, the policies are aimed at promoting employment and/or wage growth among this population, rather than just providing income support.

Table 1 describes the set of policies we consider. This set includes: (a) classroom training (CT) consisting of basic education to remedy deficiencies in general skills or vocational training to provide the skills necessary for particular jobs; (b) subsidized employment with public or private employers (WE), which includes public service

Table 1
A classification of government employment and training programs

Classroom training	
Basic education	Provides remedial general education, usually with the goal of high school certification
Classroom training in occupational skills	Provides general skills for a specific occupation or industry; duration usually less than 17 weeks
Wage and employment subsidies	
Wage and employment subsidies to private firms	Provides payments to firms, either as a lump sum per employee or as a fraction of employee wages, for hiring new workers; usually targeted at specific groups
Temporary work experience in the public or non-profit sector	Provides general work skills to youth and economically disadvantaged persons with little past employment
Public service employment	Provides temporary public sector jobs to the unemployed, especially the longterm unemployed
On-the-job training	Provides subsidies to employers to hire and train members of specific groups; when subsidy ends after 3–12 months, the employer may retain the trainee as a regular employee; training content varies from little to some; sometimes coordinated with classroom training
Job search assistance	
Employment service	Provides information on job vacancies and assists in matching workers to jobs
Job readiness training	Provides career counseling, assessment and testing to determine job readiness and to indicate appropriate search strategies; may also recommend training
Job search training and subsidies	Provides counseling, instruction in job search skills and resume preparation, job clubs, and resources such as job listings and free phones to call employers

employment (wholly subsidized temporary government jobs) and work experience (subsidized entry-level jobs at public or non-profit employers designed to introduce young people to the world of work) as well as wage supplements and fixed payments to private firms for hiring new workers; (c) subsidies to private firms for the provision of on-the-job training (OJT); (d) training in how to obtain a job; and (e) in-kind subsidies to job search such as referrals to employers and free access to job listings. Policies (d) and (e) fall under the general heading of job search assistance (JSA), which also includes the job matching services provided by the US Employment Service and similar agencies in other countries.

As we argue in more detail below, distinguishing the types of training provided is important for two reasons. First, different types of training often imply different economic models of training participation and impact and therefore different econometric estimation strategies. Second, because most existing training programs provide a mix of these services, heterogeneity in the impact of training becomes an important practical concern. As we show in Section 7, this heterogeneity has important implications for the choice of econometric methods for evaluating active labor market policies.

We do not analyze privately supplied job training despite its greater quantitative importance to modern economies (see Mincer, 1962, 1993; Heckman et al., 1997b). For example, in the United States, Mincer has estimated that such training amounts to approximately 4–5% of GDP, annually. Despite the magnitude of this investment there are surprisingly few publicly available studies of the returns to private job training, and many of those that are available do not control convincingly for the non-random allocation of training among private sector workers. Governments demand publicly justified evaluations of training programs while private firms, to the extent that they formally evaluate their training programs, keep their findings to themselves. An emphasis on objective publicly accessible evaluations is a distinctive feature of the modern welfare state, especially in an era of limited funds and public demands for accountability.

Table 2 presents the amount spent on active labor market policies by a number of OECD countries. Most OECD countries provide some mix of the employment and training services described in Table 1. Differences among countries include the relative emphasis on each type of service, the particular populations targeted for service, the total resources spent on the programs, how resources are allocated among programs and the extent to which employment and training services are integrated with other programs such as unemployment insurance or social assistance. In addition, although the programs we study are funded by governments, they are not always conducted by governments, especially in the US and the UK. In decentralized training systems, private firms and local organizations play an important role in providing employment and training services.

Table 2 reveals that many OECD countries spend substantial sums on active labor market policies. In nearly all countries, total expenditures are more than one-third of total expenditures on unemployment benefits, and some countries' expenditures on active labor market policies exceed those on unemployment benefits. Usually only a fraction of these expenditures are for CT. Further, even in countries that emphasize classroom training, governments spend substantial sums on other active labor market policies. Denmark

Table 2
Expenditures on employment and training programs in selected OECD countries as a percentage of GDP, 1994–1995[a]

Country	Adult JSA (%)	Adult CT (%)	Adult OJT (%)	Adult WE (%)	Youth All (%)	Total (%)	Disabled All (%)	Total w/disabled (%)	Income support (%)
Australia	0.20	0.17	0.09	0.13	0.07	0.66	0.07	0.73	1.64
Austria	0.13	0.12	0.02	0.03	0.01	0.31	0.06	0.37	1.44
Canada	0.20	0.34	0.00	0.02	0.02	0.60	0.00	0.60	1.54
Denmark	0.12	1.00	0.12	0.46	0.16	1.86	0.46	2.32	4.56
France	0.16	0.44	0.08	0.13	0.27	1.09	0.08	1.17	1.95
Germany	0.23	0.38	0.09	0.31	0.06	1.07	0.26	1.33	2.14
Ireland	0.14	0.48	0.03	0.25	0.43	1.33	0.15	1.48	3.25
Italy	0.08	0.02	–	–	0.83	0.93	–	0.93	1.03
Japan	0.03	0.03	0.05	–	–	0.11	–	0.11	0.35
Netherlands	0.17	0.16	0.01	0.09	0.09	0.52	0.54	1.06	3.06
Norway	0.17	0.23	0.09	0.14	0.08	0.71	0.64	1.35	1.10
Sweden	0.27	0.78	0.36	0.54	0.23	2.18	0.82	3.00	2.54
United Kingdom	0.21	0.13	0.02	0.01	0.13	0.50	0.03	0.53	1.14
United States	0.07	0.04	0.01	0.01	0.03	0.16	0.04	0.20	0.35

[a] Source: OECD (1996, Table T, pp. 206–212). Figures for Ireland and Italy are from 1991 and 1992, respectively. JSA is defined as public employment services and administration. OJT is defined as subsidies to regular employment in the private sector, or support of unemployed persons starting enterprises; WE is defined as direct job creation in the public or non-profit sector. Youth includes measures for unemployed and disadvantaged youth and support of apprenticeship and related general youth training. Income support includes unemployment compensation and early retirement benefits for labor market reasons.

spends 1% of its GDP on CT for adults, the most of any OECD country. However, this expenditure amounts to only 40% of its total spending on active labor market programs. Only in Canada is the fraction spent on CT larger. At the opposite extreme, Japan and the US spend only 0.03% and 0.04% of their GDP, respectively, on CT. However, as the table shows, these two countries also spend the smallest share of GDP on active labor market policies.

The low percentage of GDP spent on active labor market programs in the US has led some researchers to comment on the irony that despite these low expenditures, US programs have been evaluated more extensively and over a longer period of time than programs elsewhere (Haveman and Saks, 1985; Björklund, 1993). Indeed, much of what is known about the impacts of these programs and many of the methodological developments associated with evaluating them come from US evaluations.[1]

We now consider in detail each type of employment and training service in Table 1. This discussion motivates the consideration of alternative economic models of program participation and impact in Sections 6 and 7, and our focus on heterogeneity in program impacts. It also provides a context for the empirical literature on the impact of these programs that we review in Section 10.

The first category listed in Table 1 is classroom training. In many countries, CT represents the largest fraction of government expenditures on active labor market policy, and most of that expenditure is devoted to vocational training. Even in the US, where remedial programs aimed at high school dropouts and other low-skill individuals play a larger role than elsewhere, most CT programs provide vocational training. By design, most CT programs in the OECD are of limited duration. For example in Denmark, CT typically lasts 2–4 weeks (Jensen et al., 1993) while in Sweden a duration of 4 months and in the United Kingdom and the United States 3 months is more typical. Per capita expenditures on such training vary substantially, with a training slot costing approximately $7500 in Sweden and between $2000 and $3000 in the United States.[2] The Swedish figures include stipends for participants while the US figures do not.

An important difference among OECD countries that provide CT is the extent to which the training is relatively standardized and therefore less tailored to the requirements of firms or the market in general. In the 1980s and early 1990s, the Nordic countries usually provided CT in government training centers that used standardized materials and teaching methods. However, the emphasis has shifted recently, especially in Sweden, toward decentralized and firm-based training. In the United Kingdom and the US, the provision of CT is highly decentralized and its content depends on the choices made by local

[1] However, the level of total expenditure in the US is still quite large. Relative total expenditures on active labor market policies can be inferred from Table 2 using the relative sizes of each economy compared with the US. For example, the German economy is somewhat less than one-fourth the size of the US economy, and the French, Italian and British economies are approximately one-sixth the size of the US economy. Accordingly, training expenditures are somewhat greater in Germany and France, about the same in Italy, and less in the United Kingdom than in the US (see OECD, 1996, Table 1.1, p. 2).

[2] Unless otherwise indicated all monetary units are expressed in 1997 US dollars.

councils of business, political, and labor leaders. The local councils receive funding from the federal government and then subcontract for CT with private vocational and proprietary schools and local community colleges. Due to this highly decentralized structure, both participant characteristics and training content can vary substantially among locales, which suggests that the impact of training is likely to vary substantially across individuals in evaluations of such programs.

The second category of services listed in Table 1 is wage and employment subsidies. This category encompasses several different specific services which we group together due to their analytic similarity. The simplest example of this type of policy provides subsidies to private firms for hiring workers in particular groups. These subsidies may take the form of a fixed amount for each new employee hired or some fraction of the employee's wage for a period of time. In the US, the Targeted Jobs Tax Credit is an example of this type of program. Heckman et al. (1997b) discuss the empirical evidence on the effectiveness of wage and employment subsidies in greater detail.

Temporary work experience (WE) usually targets low-skilled youth or adults with poor employment histories and provides them with a job lasting 3–12 months in the public or non-profit sector. The idea of these programs is to ease the transition of these groups into regular jobs, by helping them learn about the world of work and develop good work habits. Such programs constitute a very small proportion of US training initiatives, but substantial fractions of services provided to youth in countries such as France (TUC) and the United Kingdom (Community Programmes). In public sector employment (PSE) programs, governments create temporary public sector jobs. These jobs usually require some amount of skill and are aimed at unemployed adults with recent work experience rather than youth or the disadvantaged. Except for a brief period during the late 1970s, they have not been used in the United States since the Depression era. However, they have been and remain an important component of active labor market policy in several European countries.

The third category in Table 1 is subsidized on-the-job training at private firms. The goal of subsidized OJT programs is to induce employers to provide job-relevant skills, including firm-specific skills, to disadvantaged workers. In the US, employers receive a 50% wage subsidy for up to 6 months; in the UK employers receive a lump sum per week (O'Higgins, 1994). Although evidence is limited and firm training is difficult to measure, there is a widespread view that these programs in fact provide little training, even informal on-the-job training, and are better characterized as work experience or wage subsidy programs (e.g., Breen, 1988; Hutchinson and Church, 1989).[3] Survey responses by employers who have hired or sponsored OJT trainees suggest that they value the program for its help in reducing the costs associated with hiring and retaining suitable employees more than for the opportunity to increase the skills of new workers (Begg et al., 1991).

[3] The provision of subsidized OJT is particularly hard to monitor both because on-the-job training has proven difficult to measure with survey methods (Barron et al., 1997) and because trainees often do not perceive that they have been treated any differently than their co-workers who are not subsidized. In fact, both groups may have received substantial amounts of informal on-the-job training. For evidence of the importance of informal on-the-job training in the US, see Barron et al. (1989).

For purposes of evaluation, it is almost always impossible to distinguish those OJT experiences from which new skills were acquired from those that amounted to work experience or wage subsidy without a training component. In addition, because OJT is provided by individual employers, this indeterminacy is not simply a program-specific feature, but holds among individuals within the same program. Consequently, OJT programs will likely have heterogeneous effects, and the impact, if any, of these programs will result from some combination of learning by doing, the usual training provided by the firm to new workers, and incremental training beyond that provided to unsubsidized workers.

The fourth category of services in Table 1 is job search assistance. The purpose of these services is to facilitate the matching process between workers and firms both by reducing time unemployed and by increasing match quality. The programs are usually operated by the national or local employment service, but sometimes may be subcontracted out to third parties. Included under this category are direct placement in vacant jobs, employer referrals, in-kind subsidies to search such as free access to job listings and telephones for contacting employers, career counseling, and instruction in job search skills. The last of these, which often includes instruction in general social skills, was developed in the US, but is now used in the UK, Sweden, and recently France (Björklund and Regner, 1996, p. 24). In recent years, JSA has become more popular due to its low cost, usually just a few hundred dollars per participant, and relatively solid record of performance (which we discuss in detail in Section 10).

To conclude this section, we discuss five features of employment and training programs that should be kept in mind when evaluating them. First, as the operation of these programs has become more decentralized in OECD countries, differences have emerged between how these programs were designed and how they are implemented (Hollister and Freedman, 1988). Actual practice can deviate substantially from explicit written policy.[4] Therefore, the evaluator must be careful to characterize the program as implemented when assessing its impacts.

Second, participants often receive services from more than one category in Table 1. For example, classroom training in vocational skills might be followed by job search assistance. In the UK, the Youth Training Scheme (now Youth Training) was explicitly designed to combine OJT with 13 weeks of CT. Some expensive programs combine several of the services listed in Table 1 into a single package. For example, in the US the Job Corps program for youth combines classroom training with work experience and job search assistance in a residential setting at a current cost of around $19,000 per participant. Many available survey datasets do not identify all the services received by a participant. In this case, the practice of combining together various types of training, particularly when combinations are tailored to the needs of individual trainees as in the US JTPA program, constitutes another source of heterogeneity in the impact of training. Even when administrative data are available that identify the services received, isolating the impact of particular

[4] For example, see Breen (1988) and Hollister and Freedman (1990) describing the implementation of WEP in Ireland and Hollister and Freedman (1990) and Leigh (1995) describing the implementation of JTPA in the United States.

individual services often proves difficult or impossible in practice due to the small samples receiving particular combinations of services or due to difficulties in determining the process by which individuals come to receive particular service combinations.

Third, certain features of active labor market programs affect individuals' decisions to participate in training. In some countries, such as Sweden and the United Kingdom, participation in training is a condition for receiving unemployment benefits rather than less generous social assistance payments. In the US, participation is sometimes required by a court order in lieu of alternative punishment.

Fourth, program administrators often have considerable discretion over whom they admit into government training programs. This discretion results from the fact that the number of applicants often exceeds the number of available training positions. It has long been a feature of US programs, but also has characterized programs in Austria, Denmark, Germany, Norway, and the United Kingdom (Westergard-Nielsen, 1993; Björklund and Regner, 1996; Kraus et al., 1997). Consequently, when modeling participation in training, it may be important to account for not only individual incentives, but also those of the program operators. In Section 6, we discuss the incentives facing program operators and how they affect the characteristics of participants in government training programs.

Finally, the different types of services require different economic models of program participation and impact. For example, the standard human capital model captures the essence of individual decisions to invest in vocational skills (CT). It provides little guidance to behavior regarding job search assistance or wage subsidies. In Section 6 we present economic models that describe participation in alterative programs and discuss their implications for evaluation research.

3. The evaluation problem and the parameters of interest in evaluating social programs

3.1. The evaluation problem

Constructing counterfactuals is the central problem in the literature on evaluating social programs. In the simplest form of the evaluation problem, persons are imagined as being able to occupy one of two mutually exclusive states: "0" for the untreated state and "1" for the treated state, where $D = 1$ denotes treatment and $D = 0$ denotes non-treatment. Treatment is associated with participation in the program being evaluated.[5] Associated with each state is an outcome, or set of outcomes. It is easiest to think of each state as consisting of only a single outcome measure, such as earnings, but just as easily, we can use the framework to model vectors of outcomes such as earnings, employment and

[5] In this chapter, we only consider a two potential state model in order to focus on the main ideas. Heckman (1998a) develops a multiple state model of potential outcomes for a large number of mutually exclusive states. The basic ideas in his work are captured in the two outcome models we present here.

participation in welfare programs. In the models presented in Section 6, we study an entire vector of earnings or employment at each age that result from program participation.

We can express these outcomes as a function of conditioning variables, X. Denote the potential outcomes by Y_0 and Y_1, corresponding to the untreated and treated states. Each person has a (Y_0,Y_1) pair. Assuming that means exist, we may write the (vector) of outcomes in each state as

$$Y_0 = \mu_0(X) + U_0, \tag{3.1a}$$

$$Y_1 = \mu_1(X) + U_1, \tag{3.1b}$$

where $\mathrm{E}(Y_0 \mid X) = \mu_0(X)$ and $\mathrm{E}(Y_1 \mid X) = \mu_1(X)$. To simplify the notation, we keep the conditioning on X implicit unless it serves to clarify the exposition by making it explicit. The potential outcome actually realized depends on decisions made by individuals, firms, families or government bureaucrats. This model of potential outcomes is variously attributed to Fisher (1935), Neyman (1935), Roy (1951), Quandt (1972, 1988) or Rubin (1974).

To focus on main ideas, throughout most of this chapter we assume $\mathrm{E}(U_1 \mid X) = \mathrm{E}(U_0 \mid X) = 0$, although as we note at several places in this paper, this is not strictly required. These conditions do *not* imply that $\mathrm{E}(U_1 - U_0 \mid X, D = 1) = 0$. D may depend on U_1, U_0 or $U_1 - U_0$ and X. For many of the estimators that we consider in this chapter we allow for the more general case

$$Y_0 = g_0(X) + U_0, \qquad Y_1 = g_1(X) + U_1,$$

where $\mathrm{E}(U_0 \mid X) \neq 0$ and $\mathrm{E}(U_1 \mid X) \neq 0$. Then $\mu_0(X) = g_0(X) + \mathrm{E}(U_0 \mid X)$ and $\mu_1(X) = g_1(X) + \mathrm{E}(U_1 \mid X)$.[6] Thus X is not necessarily exogenous in the ordinary econometric usage of that term.

Note also that Y may be a vector of outcomes or a time series of potential outcomes: (Y_{0t},Y_{1t}), for $t = 1,\ldots,T$, on the same type of variable. We will encounter the latter case when we analyze panel data on outcomes. In this case, there is usually a companion set of X variables which we will sometimes assume to be strictly exogenous in the conventional econometric meaning of that term: $\mathrm{E}(U_{0t} \mid X) = 0$, $\mathrm{E}(U_{1t} \mid X) = 0$ where $X = (X_1,\ldots,X_T)$. In defining a sequence of "treatment on the treated" parameters, $\mathrm{E}(Y_{1t} - Y_{0t} \mid X, D = 1)$, $t = 1,\ldots,T$, this assumption allows us to abstract from any dependence between U_{1t}, U_{0t} and X. It excludes differences in U_{1t} and U_{0t} arising from X dependence and allows us to focus on differences in outcomes solely attributable to D. While convenient, this assumption is overly strong.

However, we stress that the exogeneity assumption in either cross-section or panel contexts is only a matter of convenience and is not strictly required. What is required for an interpretable definition of the "treatment on the treated" parameter is avoiding conditioning on X variables *caused* by D even holding $Y^{\mathrm{P}} = ((Y_{01},Y_{11}),\ldots,(Y_{0T},Y_{1T}))$ fixed

[6] For example, an exogeneity assumption is not required when using social experiments to identify $\mathrm{E}(Y_1 - Y_0 \mid X, D = 1)$.

where Y^P is the vector of potential outcomes. More precisely, we require that for the conditional density of the data

$$f(X \mid D, Y^P) = f(X \mid Y^P),$$

i.e., we require that the realization of D does not determine X given the vector of potential outcomes. Otherwise, the parameter $E(Y_1 - Y_0 \mid X, D = 1)$ does not capture the full effect of treatment on the treated as it operates through all channels and certain other technical problems discussed in Heckman (1998a) arise. In order to obtain $E(Y_{1t} - Y_{0t} \mid X, D = 1)$ defined on subsets of X, say X_c, simply integrate out $E(Y_{1t} - Y_{0t} \mid X, D)$ against the density $f(\tilde{X}_c \mid D = 1)$ where $\tilde{X}_c$ is the portion of X not in $X_c : X = (X_c, \tilde{X}_c)$.

Note, finally, that the choice of a base state "0" is arbitrary. Clearly the roles of "0" and "1" can be reversed. In the case of human capital investments, there is a natural base state. But for many other evaluation problems the choice of a base is arbitrary. Assumptions appropriate for one choice of "0" and "1" need not carry over to the opposite choice. With this cautionary note in mind, we proceed as if a well-defined base state exists.

In many problems it is convenient to think of "0" as a benchmark "no treatment" state. The gain to the individual of moving from "0" to "1" is given by

$$\Delta = Y_1 - Y_0. \tag{3.2}$$

If one could observe both Y_0 and Y_1 for the same person at the same time, the gain Δ would be known for each person. The fundamental evaluation problem arises because we do not know both coordinates of (Y_1, Y_0) and hence Δ for anybody. All approaches to solving this problem attempt to estimate the missing data. These attempts to solve the evaluation problem differ in the assumptions they make about how the missing data are related to the available data, and what data are available. Most approaches to evaluation in the social sciences accept the impossibility of constructing Δ for anyone. Instead, the evaluation problem is redefined from the individual level to the population level to estimate the mean of Δ, or some other aspect of the distribution of Δ, for various populations of interest. The question becomes what features of the distribution of Δ should be of interest and for what populations should it be defined?

3.2. The counterfactuals of interest

There are many possible counterfactuals of interest for evaluating a social program. One might like to compare the state of the world in the presence of the program to the state of the world if the program were operated in a different way, or to the state of the world if the program did not exist at all, or to the state of the world if alternative programs were used to replace the present program. A full evaluation entails an enumeration of all outcomes of interest for all persons both in the current state of the world and in all the alternative states of interest, and a mechanism for valuing the outcomes in the different states.

Outcomes of interest in program evaluations include the direct benefits received, the level of behavioral variables for participants and non-participants and the payments for the

program, for both participants and non-participants, including taxes levied to finance a publicly provided program. These measures would be displayed for each individual in the economy to characterize each state of the world.

In a Robinson Crusoe economy, participation in a program is a well-defined event. In a modern economy, almost everyone participates in each social program either directly or indirectly. A training program affects more than the trainees. It also affects the persons with whom the trainees compete in the labor market, the firms that hire them and the taxpayers who finance the program. The impact of the program depends on the number and composition of the trainees. Participation in a program does not mean the same thing for all people.

The traditional evaluation literature usually defines the effect of participation to be the effect of the program on participants explicitly enrolled in the program. These are the *Direct Effects*. They exclude the effects of a program that do not flow from direct participation, known as the *Indirect Effects*. This distinction appears in the pioneering work of H.G. Lewis on measuring union relative wage effects (Lewis, 1963). His insights apply more generally to all evaluation problems in social settings.

There may be indirect effects for both participants and non-participants. Thus a participant may pay taxes to support the program just as persons who do not participate may also pay taxes. A firm may be an indirect beneficiary of the lower wages resulting from an expansion of the trained workforce. The conventional econometric and statistical literature ignores the indirect effects of programs and equates "treatment" outcomes with the direct outcome Y_1 in the program state and "no treatment" with the direct outcome Y_0 in the no program state.

Determining all outcomes in all states is not enough to evaluate a program. Another aspect of the evaluation problem is the valuation of the outcomes. In a democratic society, aggregation of the evaluations and the outcomes in a form useful for social deliberations also is required. Different persons may value the same state of the world differently even if they experience the same "objective" outcomes and pay the same taxes. Preferences may be interdependent. Redistributive programs exist, in part, because of altruistic or paternalistic preferences. Persons may value the outcomes of other persons either positively or negatively. Only if one person's preferences are dominant (the idealized case of a social planner with a social welfare function) is there a unique evaluation of the outcomes associated with each possible state from each possible program.

The traditional program evaluation literature assumes that the valuation of the direct effects of the program boils down to the effect of the program on GDP. This assumption ignores the important point that different persons value the same outcomes differently and that the democratic political process often entails coalitions of persons who value outcomes in different ways. Both efficiency and equity considerations may receive different weights from different groups. Different mechanisms for aggregating evaluations and resolving social conflicts exist in different societies. Different types of information are required to evaluate a program under different modes of social decision making.

Both for pragmatic and political reasons, government social planners, statisticians or policy makers may value objective output measures differently than the persons or institu-

tions being evaluated. The classic example is the value of non-market time (Greenberg, 1997). Traditional program evaluations exclude such valuations largely because of the difficulty of imputing the value and quantity of non-market time. By doing this, however, these evaluations value labor supply in the market sector at the market wage, but value labor supply in the non-market sector at a zero wage. By contrast, individuals value labor supply in the non-market sector at their reservation wage. In this example, two different sets of preferences value the same outcomes differently. In evaluating a social program in a society that places weight on individual preferences, it is appropriate to recognize personal evaluations and that the same outcome may be valued in different ways by different social actors.

Programs that embody redistributive objectives inherently involve different groups. Even if the taxpayers and the recipients of the benefits of a program have the same preferences, their valuations of a program will, in general, differ. Altruistic considerations often motivate such programs. These often entail private valuations of *distributions* of program *impacts* – how much recipients gain over what they would experience in the absence of the program (see Heckman and Smith, 1993, 1995, 1998a; Heckman et al., 1997c).

Answers to many important evaluation questions require knowledge of the distribution of program gains especially for programs that have a redistributive objective or programs for which altruistic motivations play a role in motivating the existence of the program. Let $D = 1$ denote direct participation in the program and $D = 0$ denote direct non-participation. To simplify the argument in this section, ignore any indirect effects. From the standpoint of a detached observer of a social program who takes the base state values (denoted "0") as those that would prevail in the absence of the program, it is of interest to know, among other things,

(A) the proportion of people taking the program who benefit from it:

$$\Pr(Y_1 > Y_0 \mid D = 1) = \Pr(\Delta > 0 \mid D = 1);$$

(B) the proportion of the total population benefiting from the program:

$$\Pr(Y_1 > Y_0 \mid D = 1)\Pr(D = 1) = \Pr(\Delta > 0 \mid D = 1)\Pr(D = 1);$$

(C) selected quantiles of the impact distribution:

$$\inf_{\Delta} \{\Delta : F(\Delta \mid D = 1) > q\},$$

where q is a quantile of the distribution and "inf" is the smallest attainable value of Δ that satisfies the condition stated in the braces;

(D) the distribution of gains at selected base state values:

$$F(\Delta \mid D = 1, Y_0 = y_0);$$

(E) the increase in the proportion of outcomes above a certain threshold $\bar{y}$ due to a policy:

$$\Pr(Y_1 > \bar{y} \mid D = 1) - \Pr(Y_0 > \bar{y} \mid D = 1).$$

Measure (A) is of interest in determining how widely program *gains* are distributed among participants. Participants in the political process with preferences over distributions of program outcomes would be unlikely to assign the same weight to two programs with the same mean outcome, one of which produced favorable outcomes for only a few persons while the other distributed gains more broadly. When considering a program, it is of interest to determine the proportion of participants who are harmed as a result of program participation, indicated by $\Pr(Y_1 < Y_0 \mid D = 1)$. Negative mean impact results might be acceptable if most participants gain from the program. These features of the outcome distribution are likely to be of interest to evaluators even if the persons studied do not know their Y_0 and Y_1 values in advance of participating in the program.

Measure (B) is the proportion of the entire population that benefits from the program, assuming that the costs of financing the program are broadly distributed and are not perceived to be related to the specific program being evaluated. If voters have correct expectations about the joint distribution of outcomes, it is of interest to politicians to determine how widely program benefits are distributed. At the same time, large program gains received by a few persons may make it easier to organize interest groups in support of a program than if the same gains are distributed more widely.

Evaluators interested in the distribution of program benefits would be interested in measure (C). Evaluators who take a special interest in the impact of a program on recipients in the lower tail of the base state distribution would find measure (D) of interest. It reveals how the distribution of gains depends on the base state for participants. Measure (E) provides the answer to the question "does the distribution of outcomes for the participants dominate the distribution of outcomes if they did not participate?" (see Heckman et al., 1997c; Heckman and Smith, 1998a). Expanding the scope of the discussion to evaluate the indirect effects of the program makes it more likely that estimating distributional impacts plays an important part in conducting program evaluations.

3.3. The counterfactuals most commonly estimated in the literature

The evaluation problem in its most general form for distributions of outcomes is formidable and is not considered in depth either in this chapter or in the literature (Heckman et al., 1997c; Heckman and Smith, 1998a, consider identification and estimation of counterfactual distributions). Instead, in this chapter we focus on counterfactual means, and consider a form of the problem in which analysts have access to information on persons who are in one state or the other at any time, and for certain time periods there are some persons in both states, but there is no information on any single person who is in both states at the same time. As discussed in Heckman (1998a) and Heckman and Smith (1998a), a crucial assumption in the traditional evaluation literature is that the no treatment state approximates the no program state. This would be true if indirect effects are negligible.

Most of the empirical work in the literature on evaluating government training programs focuses on means and in particular on one mean counterfactual: the mean direct effect of

treatment on those who take treatment. The transition from the individual to the group level counterfactual recognizes the inherent impossibility of observing the same person in both states at the same time. By dealing with aggregates, rather than individuals, it is sometimes possible to estimate group impact measures even though it may be impossible to measure the impacts of a program on any particular individual. To see this point more formally, consider the switching regression model with two regimes denoted by "1" and "0" (Quandt, 1972). The observed outcome Y is given by

$$Y = DY_1 + (1 - D)Y_0. \tag{3.3}$$

When $D = 1$ we observe Y_1; when $D = 0$ we observe Y_0.

To cast the foregoing model in a more familiar-looking form, and to distinguish it from conventional regression models, express the means in (3.1a) and (3.1b) in more familiar linear regression form:

$$\mathrm{E}(Y_j \mid X) = \mu_j(X) = X\beta_j, \qquad j = 0, 1.$$

With these expressions, substitute from (3.1a) and (3.1b) into (3.3) to obtain

$$Y = D(\mu_1(X) + U_1) + (1 - D)(\mu_0(X) + U_0).$$

Rewriting,

$$Y = \mu_0(X) + D(\mu_1(X) - \mu_0(X) + U_1 - U_0) + U_0.$$

Using the linear regression representation, we obtain

$$Y = X\beta_0 + D(X(\beta_1 - \beta_0) + U_1 - U_0) + U_0. \tag{3.4}$$

Observe that from the definition of a conditional mean, $\mathrm{E}(U_0 \mid X) = 0$ and $\mathrm{E}(U_1 \mid X) = 0$.

The parameter most commonly invoked in the program evaluation literature, although not the one actually estimated in social experiments or in most non-experimental evaluations, is the effect of randomly picking a person with characteristics X and moving that person from "0" to "1":

$$\mathrm{E}(Y_1 - Y_0 \mid X) = \mathrm{E}(\Delta \mid X).$$

In terms of the switching regression model this parameter is the coefficient on D in the non-error component of the following "regression" equation:

$$Y = \mu_0(X) + D(\mu_1(X) - \mu_0(X)) + \{U_0 + D(U_1 - U_0)\}$$

$$= \mu_0(X) + D(\mathrm{E}(\Delta \mid X)) + \{U_0 + D(U_1 - U_0)\}$$

$$= X\beta_0 + DX(\beta_1 - \beta_0) + \{U_0 + D(U_1 - U_0)\}, \tag{3.5}$$

where the term in braces is the "error."

If the model is specialized so that there are K regressors plus an intercept and $\beta_1 = (\beta_{10},\ldots,\beta_{1K})$ and $\beta_0 = (\beta_{00},\ldots,\beta_{0K})$, where the intercepts occupy the first position, and the

slope coefficients are the same in both regimes:

$$\beta_{1j} = \beta_{0j} = \beta_j, \qquad j = 1, \ldots, K$$

and $\beta_{00} = \beta_0$ and $\beta_{10} - \beta_{00} = \alpha$, the parameter under consideration reduces to α:

$$\mathrm{E}(Y_1 - Y_0 \mid X) = \beta_{10} - \beta_{00} = \alpha. \tag{3.6}$$

The regression model for this special case may be written as

$$Y = X\beta + D\alpha + \{U_0 + D(U_1 - U_0)\}. \tag{3.7}$$

It is non-standard from the standpoint of elementary econometrics because the error term has a component that switches on or off with D. In general, its mean is not zero because $\mathrm{E}[U_0 + D(U_1 - U_0)] = \mathrm{E}(U_1 - U_0 \mid D = 1)\Pr(D = 1)$. If $U_1 - U_0$, or variables statistically dependent on it, help determine D, $\mathrm{E}(U_1 - U_0 \mid D = 1) \neq 0$. Intuitively, if persons who have high gains $(U_1 - U_0)$ are more likely to appear in the program, then this term is positive.

In practice most non-experimental and experimental studies do not estimate $\mathrm{E}(\Delta \mid X)$. Instead, most non-experimental studies estimate the effect of treatment on the treated, $\mathrm{E}(\Delta \mid X, D = 1)$. This parameter conditions on participation in the program as follows:

$$\mathrm{E}(\Delta \mid X, D = 1) = \mathrm{E}(Y_1 - Y_0 \mid X, D = 1) = X(\beta_1 - \beta_0) + \mathrm{E}(U_1 - U_0 \mid X, D = 1). \tag{3.8}$$

It is the coefficient on D in the non-error component of the following regression equation:

$$\begin{aligned} Y &= \mu_0(X) + D[\mathrm{E}(\Delta \mid X, D = 1)] + \{U_0 + D[(U_1 - U_0) - \mathrm{E}(U_1 - U_0 \mid X, D = 1)]\} \\ &= X\beta_0 + D[X(\beta_1 - \beta_0) + \mathrm{E}(U_1 - U_0 \mid X, D = 1)] \\ &\quad + \{U_0 + D[(U_1 - U_0) - \mathrm{E}(U_1 - U_0 \mid X, D = 1)]\}. \end{aligned} \tag{3.9}$$

$\mathrm{E}(\Delta \mid X, D = 1)$ is a non-standard parameter in conventional econometrics. It combines "structural" parameters $(X(\beta_1 - \beta_0))$ with the means of the unobservables $(\mathrm{E}(U_1 - U_0 \mid X, D = 1))$. It measures the average gain in the outcome for persons who choose to participate in a program compared to what they would have experienced in the base state. It computes the average gain in terms of both observables and unobservables. It is the latter that makes the parameter look non-standard. Most econometric activity is devoted to separating β_0 and β_1 from the effects of the regressors on U_1 and U_0. Parameter (3.8) combines these effects.

This parameter is implicitly defined conditional on the current levels of participation in the program in society at large. Thus it recognizes social interaction. But at any point in time the aggregate participation level is just a single number, and the composition of trainees is fixed. From a single cross-section of data, it is not possible to estimate how variation in the levels and composition of participants in a program affect the parameter.

The two evaluation parameters we have just presented are the same if we assume that $U_1 - U_0 = 0$, so the unobservables are common across the two states. From (3.9) we now

have $Y_1 - Y_0 = \mu_1(X) - \mu_0(X) = X(\beta_1 - \beta_0)$. The difference between potential outcomes in the two states is a function of X but not of unobservables. Further specializing the model to one of intercept differences (i.e., $Y_1 - Y_0 = \alpha$), requires that the difference between potential outcomes is a constant. The associated regression can be written as the familiar-looking dummy variable regression model:

$$Y = X\beta + D\alpha + U, \tag{3.10}$$

where $\mathrm{E}(U) = 0$. The parameter α is easy to interpret as a standard structural parameter and the specification (3.10) looks conventional. In fact, model (3.10) dominates the conventional evaluation literature. The validity of many conventional instrumental variables methods and longitudinal estimation strategies is contingent on this specification as we document below. The conventional econometric evaluation literature focuses on α, or more rarely, $X(\beta_1 - \beta_0)$, and the selection problem arises from the correlation between D and U.

While familiar, the framework of (3.10) is very special. Potential outcomes (Y_1, Y_0) differ only by a constant $(Y_1 - Y_0 = \alpha)$. The best Y_1 is the best Y_0. All people gain or lose the same amount in going from "0" to "1". There is no heterogeneity in gains. Even in the more general case, with $\mu_1(X)$ and $\mu_0(X)$ distinct, or $\beta_1 \neq \beta_0$ in the linear regression representation, so long as $U_1 = U_0$ among people with the same X, there is no heterogeneity in the outcomes moving from "0" to "1". This assumed absence of heterogeneity in response to treatments is strong. When tested, it is almost always rejected (see Heckman et al., 1997c, and the evidence presented below).

There is one case when $U_1 \neq U_0$, where the two parameters of interest are still equal even though there is dispersion in gain Δ. This case occurs when

$$\mathrm{E}(U_1 - U_0 \mid X, D = 1) = 0. \tag{3.11}$$

Condition (3.11) arises when conditional on X, D does not explain or predict $U_1 - U_0$. This condition could arise if agents who select into state "1" from "0" either do not know or do not act on $U_1 - U_0$, or information dependent on $U_1 - U_0$, in making their decision to participate in the program. Ex post, there is heterogeneity, but ex ante it is not acted on in determining participation in the program.

When the gain does not affect individuals' decisions to participate in the program, the error terms (the terms in braces in (3.7) and (3.9)) have conventional properties. The only bias in estimating the coefficients on D in the regression models arises from the dependence between U_0 and D, just as the only source of bias in the common coefficient model is the covariance between U and D when $\mathrm{E}(U \mid X) = 0$. To see this point take the expectation of the terms in braces in (3.7) and (3.9), respectively, to obtain the following:

$$\mathrm{E}(U_0 + D(U_1 - U_0) \mid X, D) = \mathrm{E}(U_0 \mid X, D)$$

and

$$\mathrm{E}(U_0 + D[(U_1 - U_0) - \mathrm{E}(U_1 - U_0 \mid X, D = 1)] \mid X, D) = \mathrm{E}(U_0 \mid X, D).$$

A problem that remains when condition (3.11) holds is that the D component in the error

terms contributes a component of variance to the model and so makes the model heteroscedastic:

$$\mathrm{Var}(U_0 + D(U_1 - U_0) \mid X, D) = \mathrm{Var}(U_0 \mid X, D)$$

$$+2\mathrm{Cov}(U_0, U_1 - U_0 \mid X, D)D + \mathrm{Var}(U_1 - U_0 \mid X, D)D.$$

The distinction between a model with $U_1 = U_0$, and one with $U_1 \neq U_0$, is fundamental to understanding modern developments in the program evaluation literature. When $U_1 = U_0$ and we condition on X, *everyone* with the same X has the same treatment effect. The evaluation problem greatly simplifies and one parameter answers all of the conceptually distinct evaluation questions we have posed. "Treatment on the treated" is the same as the effect of taking a person at random and putting him/her into the program. The distributional questions (A)–(E) all have simple answers because everyone with the same X has the same Δ. Eq. (3.10) is amenable to analysis by conventional econometric methods. Eliminating the covariance between D and U is the central problem in this model.

When $U_1 \neq U_0$, but (3.11) characterizes the program being evaluated, most of the familiar econometric intuition remains valid. This is the "random coefficient" model with the coefficient on D "random" (from the standpoint of the observing economist), but uncorrelated with D. The central problem in this model is covariance between U_0 and D and the only additional econometric problem arises in accounting for heteroscedasticity in getting the right standard errors for the coefficients. In this case, the response to treatment varies among persons with the same X values. The mean effect of treatment on the treated and the effect of treatment on a randomly chosen person are the same.

In the general case when $U_1 \neq U_0$ and (3.11) no longer holds, we enter a new world not covered in the traditional econometric evaluation literature. A variety of different treatment effects can be defined. Conventional econometric procedures often break down or require substantial modification. The error term for the model (3.5) has a non-zero mean.[7] Both error terms are heteroscedastic. The distinctions among these three models – (a) the coefficient on D is fixed (given X) for everyone; (b) the coefficient on D is variable (given X), but does not help determine program participation; and (c) the coefficient on D is variable (given X) and does help determine program participation – are fundamental to this chapter and the entire literature on program evaluation.

3.4. Is treatment on the treated an interesting economic parameter?

What economic question does parameter (3.8) answer? How does it relate to the conventional parameter of interest in cost-benefit analysis – the effect of a program on GDP? In order to relate the parameter (3.8) with the parameters needed to perform traditional cost-benefit analysis, it is fruitful to consider a more general framework. Following our previous discussion, we consider two discrete states or sectors corresponding to direct

[7] $\mathrm{E}[U_0 + D(U_1 - U_0) \mid X] = \mathrm{E}(U_1 - U_0 \mid X, D = 1)\Pr(D = 1 \mid X) \neq 0.$

participation and non-participation and a vector of policy variables φ that affect the outcomes in both states and the allocation of all persons to states or sectors. The policy variables may be discrete or continuous. Our framework departs from the conventional treatment effect literature and allows for general equilibrium effects.

Assuming that costless lump-sum transfers are possible, that a single social welfare function governs the distribution of resources and that prices reflect true opportunity costs, traditional cost-benefit analysis (see, e.g., Harberger, 1971) seeks to determine the impact of programs on the total output of society. Efficiency becomes the paramount criterion in this framework, with the distributional aspects of policies assumed to be taken care of by lump sum transfers and taxes engineered by an enlightened social planner. In this framework, impacts on total output are the only objects of interest in evaluating programs. The distribution of program impacts is assumed to be irrelevant. This framework is favorable to the use of mean outcomes to evaluate social programs.

Within the context of the simple framework discussed in Section 3.1, let Y_1 and Y_0 be individual output which trades at a constant relative price of "1" set externally and not affected by the decisions of the agents we analyze. Alternatively, assume that the policies we consider do not alter relative prices. Let φ be a vector of policy variables which operate on all persons. These also generate indirect effects. $c(\varphi)$ is the social cost of φ denominated in "1" units. We assume that $c(0) = 0$ and that c is convex and increasing in φ. Let $N_1(\varphi)$ be the number of persons in state "1" and $N_0(\varphi)$ be the number of persons in state "0". The total output of society is

$$N_1(\varphi)\mathrm{E}(Y_1 \mid D = 1, \varphi) + N_0(\varphi)\mathrm{E}(Y_0 \mid D = 0, \varphi) - c(\varphi),$$

where $N_1(\varphi) + N_0(\varphi) = \bar{N}$ is the total number of persons in society. For simplicity, we assume that all persons have the same person-specific characteristics X. Vector φ is general enough to include financial incentive variables for participation in the program as well as mandates that assign persons to a particular state. A policy may benefit some and harm others.

Assume for convenience that the treatment choice and mean outcome functions are differentiable and for the sake of argument further assume that φ is a scalar. Then the change in output in response to a marginal increase in φ from any given position is

$$\Delta(\varphi) = \frac{\partial N_1(\varphi)}{\partial \varphi}[\mathrm{E}(Y_1 \mid D = 1, \varphi) - \mathrm{E}(Y_0 \mid D = 0, \varphi)]$$

$$+ N_1(\varphi)\left[\frac{\partial \mathrm{E}(Y_1 \mid D = 1, \varphi)}{\partial \varphi}\right] + N_0(\varphi)\left[\frac{\partial \mathrm{E}(Y_0 \mid D = 0, \varphi)}{\partial \varphi}\right] - \frac{\partial c(\varphi)}{\partial \varphi}. \tag{3.12}$$

The first term arises from the transfer of persons across sectors that is induced by the policy change. The second term arises from changes in output within each sector induced by the policy change. The third term is the marginal social cost of the change.

In principle, this measure could be estimated from time-series data on the change in aggregate GDP occurring after the program parameter φ is varied. Assuming a well-

defined social welfare function and making the additional assumption that prices are constant at initial values, an increase in GDP evaluated at base period prices raises social welfare provided that feasible bundles can be constructed from the output after the social program parameter is varied so that all losers can be compensated. (See, e.g., Laffont, 1989, p. 155, or the comprehensive discussion in Chipman and Moore, 1976).

If marginal policy changes have no effect on intra-sector mean output, the bracketed elements in the second set of terms are zero. In this case, the parameters of interest for evaluating the impact of the policy change on GDP are

(i) $\partial N_1(\varphi)/\partial\varphi$; the number of people entering or leaving state 1.

(ii) $\mathrm{E}(Y_1 \mid D = 1, \varphi) - \mathrm{E}(Y_0 \mid D = 0, \varphi)$; the mean output difference between sectors.

(iii) $\partial c(\varphi)/\partial\varphi$; the social marginal cost of the policy.

It is revealing that nowhere on this list are the parameters that receive the most attention in the econometric policy evaluation literature. (See, e.g., Heckman and Robb, 1985a). These are "the effect of treatment on the treated":

(a) $\mathrm{E}(Y_1 - Y_0 \mid D = 1, \varphi)$ or

(b) $\mathrm{E}(Y_1 \mid \varphi = \bar{\varphi}) - \mathrm{E}(Y_0 \mid \varphi = 0)$ where $\varphi = \bar{\varphi}$ sets $N_1(\bar{\varphi}) = \bar{N}$, the effect of universal coverage for the program.

Parameter (ii) can be estimated by taking simple mean differences between the outputs in the two sectors; no adjustment for selection bias is required. Parameter (i) can be obtained from knowledge of the net movement of persons across sectors in response to the policy change, something usually neglected in micro policy evaluation (for exceptions, see Moffitt, 1992; Heckman, 1992). Parameter (iii) can be obtained from cost data. Full social marginal costs should be included in the computation of this term. The typical micro evaluation neglects all three terms. Costs are rarely collected and gross outcomes are typically reported; entry effects are neglected and term (ii) is usually "adjusted" to avoid selection bias when in fact, no adjustment is needed to estimate the impact of the program on GDP.

It is informative to place additional structure on this model. This leads to a representation of a criterion that is widely used in the literature on microeconomic program evaluation and also establishes a link with the models of program participation used in the later sections of this chapter. Assume a binary choice random utility framework. Suppose that agents make choices based on net utility and that policies affect participant utility through an additively-separable term $k(\varphi)$ that is assumed scalar and differentiable. Net utility is

$$U = X + k(\varphi),$$

where k is monotonic in φ and where the joint distributions of (Y_1, X) and (Y_0, X) are $F(y_1, x)$ and $F(y_0, x)$, respectively. The underlying variables are assumed to be continuously distributed. In the special case of the Roy model of self-selection (see Heckman and Honoré, 1990, for one discussion) $X = Y_1 - Y_0$,

$$D = 1(U \geq 0) = 1(X \geq -k(\varphi)),$$

where "1" is the indicator function ($1(Z \geq 0) = 1$ if $Z \geq 0$; $= 0$ otherwise),

$$N_1(\varphi) = \bar{N}\Pr(U \geq 0) = \bar{N}\int_{-k(\varphi)}^{\infty} f(x)\mathrm{d}x,$$

and

$$N_0(\varphi) = \bar{N}\Pr(U < 0) = \bar{N}\int_{-\infty}^{-k(\varphi)} f(x)\mathrm{d}x,$$

where $f(x)$ is the density of x. Total output is

$$\bar{N}\int_{-\infty}^{\infty} y_1 \int_{-k(\varphi)}^{\infty} f(y_1, x \mid \varphi)\mathrm{d}x\,\mathrm{d}y_1 + \bar{N}\int_{-\infty}^{\infty} y_0 \int_{-\infty}^{-k(\varphi)} f(y_0, x \mid \varphi)\mathrm{d}x\,\mathrm{d}y_0 - c(\varphi).$$

Under standard conditions (see, e.g., Royden, 1968), we may differentiate this expression to obtain the following expression for the marginal change in output with respect to a change in φ:

$$\Delta(\varphi) = \bar{N}k'(\varphi)f_x(-k(\varphi))[\mathrm{E}(Y_1 \mid D = 1, x = -k(\varphi), \varphi) - \mathrm{E}(Y_0 \mid D = 0, x = -k(\varphi), \varphi)]$$

$$+\bar{N}[\int_{-\infty}^{\infty} y_1 \int_{-k(\varphi)}^{\infty} \frac{\partial f(y_1, x \mid \varphi)}{\partial \varphi}\mathrm{d}x\,\mathrm{d}y_1 + \int_{-\infty}^{\infty} y_0 \int_{-\infty}^{-k(\varphi)} \frac{\partial f(y_0, x \mid \varphi)}{\partial \varphi}\mathrm{d}x\,\mathrm{d}y_0] - \frac{\partial c(\varphi)}{\partial \varphi}. \tag{3.13}$$

This model has a well-defined margin: $X = -k(\varphi)$, which is the utility of the marginal entrant into the program. The utility of the participant might be distinguished from the objective of the social planner who seeks to maximize total output. The first set of terms corresponds to the gain arising from the movement of persons at the margin (the term in brackets) weighted by the proportion of the population at the margin, $k'(\varphi)f_x(-k(\varphi))$, times the number of people in the population. This term is the net gain from switching sectors. The expression in brackets in the first term is a limit form of the "local average treatment effect" of Imbens and Angrist (1994) which we discuss further in our discussion of instrumental variables in Section 7.4.5. The second set of terms is the intrasector change in output resulting from a policy change. This includes both direct and indirect effects. The second set of terms is ignored in most evaluation studies. It describes how people who do not switch sectors are affected by the policy. The third term is the direct marginal social cost of the policy change. It includes the cost of administering the program plus the opportunity cost of consumption foregone to raise the taxes used to finance the program. Below we demonstrate the empirical importance of accounting for the full social costs of programs.

At an optimum, $\Delta(\varphi) = 0$, provided standard second order conditions are satisfied. Marginal benefit should equal the marginal cost. We can use either a cost-based measure of marginal benefit or a benefit-based measure of cost to evaluate the marginal gains or marginal costs of the program, respectively.

Observe that the local average treatment effect is simply the effect of treatment on the treated for persons at the margin ($X = -k(\varphi)$):

$$\mathrm{E}(Y_1 \mid D = 1, X = -k(\varphi), \varphi) - \mathrm{E}(Y_0 \mid D = 0, X = -k(\varphi), \varphi)$$

$$= \mathrm{E}(Y_1 - Y_0 \mid D = 1, X = -k(\varphi), \varphi). \tag{3.14}$$

This expression is obvious once it is recognized that the set $X = -k(\varphi)$ is the indifference set. Persons in that set are indifferent between participating in the program and not participating. The Imbens and Angrist (1994) parameter is a marginal version of the "treatment on the treated" evaluation parameter for gross outcomes. This parameter is one of the ingredients required to produce an evaluation of the impact of a marginal change in the social program on total output but it ignores costs and the effect of a change in the program on the outcomes of persons who do not switch sectors.[8]

The conventional evaluation parameter,

$$\mathrm{E}(Y_1 - Y_0 \mid D = 1, x, \varphi)$$

does not incorporate costs, does not correspond to a marginal change and includes rents accruing to persons. This parameter is in general inappropriate for evaluating the effect of a policy change on GDP. However, under certain conditions which we now specify, this parameter is informative about the gross gain accruing to the economy from the existence of a program at level $\tilde{\varphi}$ compared to the alternative of shutting it down. This is the information required for an "all or nothing" evaluation of a program.

The appropriate criterion for an all or nothing evaluation of a policy at level $\varphi = \tilde{\varphi}$ is

$$A(\tilde{\varphi}) = \{N_1(\tilde{\varphi})\mathrm{E}(Y_1 \mid D = 1, \varphi = \tilde{\varphi}) + N_0(\tilde{\varphi})\mathrm{E}(Y_0 \mid D = 0, \varphi = \tilde{\varphi}) - c(\tilde{\varphi})\}$$

$$-\{N_1(0)\mathrm{E}(Y_1 \mid D = 1, \varphi = 0) + N_0(0)\mathrm{E}(Y_0 \mid D = 0, \varphi = 0)\},$$

where $\varphi = 0$ corresponds to the case where there is no program, so that $N_1(0) = 0$ and $N_0(0) = \bar{N}$. If $A(\tilde{\varphi}) > 0$, total output is increased by establishing the program at level $\tilde{\varphi}$.

In the special case where the outcome in the benchmark state "0" is the same whether or not the program exists, so

$$\mathrm{E}(Y_0 \mid D = 0, \varphi = \tilde{\varphi}) = \mathrm{E}(Y_0 \mid D = 0, \varphi = 0). \tag{3.15}$$

and

$$\mathrm{E}(Y_0 \mid D = 1, \varphi = \tilde{\varphi}) = \mathrm{E}(Y_0 \mid D = 1, \varphi = 0).$$

This condition defines the absence of general equilibrium effects in the base state so the no program state for non-participants is the same as the non-participation state. Assumption (3.15) is what enables analysts to generalize from partial equilibrium to general equili-

[8] Heckman and Smith (1998a) and Heckman (1997) present comprehensive discussions of the Imbens and Angrist (1994) parameter. We discuss this parameter further in Section 7.4.5. One important difference between their parameter and the traditional treatment on the treated parameter is that the latter excludes variables like φ from the conditioning set, but the Imbens–Angrist parameter includes them.

brium settings. Recalling that $\bar{N} = N_1(\varphi) + N_0(\varphi)$, when (3.15) holds we have[9]

$$A(\tilde{\varphi}) = N_1(\tilde{\varphi})\mathrm{E}(Y_1 - Y_0 \mid D = 1, \varphi = \tilde{\varphi}) - c(\tilde{\varphi}). \tag{3.16}$$

Given costless redistribution of the benefits, the output-maximizing solution for φ also maximizes social welfare. For this important case, which is applicable to small-scale social programs with partial participation, the measure "treatment on the treated" which we focus on in this chapter is justified. For evaluating the effect of marginal variation or "fine-tuning" of existing policies, measure $\Delta(\varphi)$ is more appropriate.[10]

4. Prototypical solutions to the evaluation problem

An evaluation entails making some comparison between "treated" and "untreated" persons. This section considers three widely used comparisons for estimating the impact of treatment on the treated: $\mathrm{E}(Y_1 - Y_0 \mid X, D = 1)$. All use some form of comparison to construct the required counterfactual $\mathrm{E}(Y_0 \mid X, D = 1)$. Data on $\mathrm{E}(Y_1 \mid X, D = 1)$ are available from program participants. A person who has participated in a program is paired with an "otherwise comparable" person or set of persons who have not participated in it. The set may contain just one person. In most applications of the method, the paired partner is not literally assumed to be a replica of the treated person in the untreated state although some panel data evaluation estimators make such an assumption. Thus, in general, $\Delta = Y_1 - Y_0$ is not estimated exactly. Instead, the outcome of the paired partners is treated as a proxy for Y_0 for the treated individual and the population mean difference between treated and untreated persons is estimated by averaging over all pairs. The method can be applied symmetrically to non-participants to estimate what they would have earned if they had participated. For that problem the challenge is to find $\mathrm{E}(Y_1 \mid X, D = 0)$ since the data on non-participants enables one to identify $\mathrm{E}(Y_0 \mid X, D = 0)$.

A major difficulty with the application of this method is providing some objective way of demonstrating that a candidate partner or set of partners is "otherwise comparable." Many econometric and statistical methods are available for adjusting differences between persons receiving treatment and potential matching partners which we discuss in Section 7.

4.1. The before–after estimator

In the empirical literature on program evaluation, the most commonly-used evaluation strategy compares a person with himself/herself. This is a comparison strategy based on longitudinal data. It exploits the intuitively appealing idea that persons can be in both states at different times, and that outcomes measured in one state at one time are good proxies for outcomes in the same state at other times at least for the no-treatment state. This gives rise

[9] Condition (3.15) is stronger than what is required to justify (3.16). The condition only has to hold for the subset of the population ($N_0(\varphi)$ in number) who would not participate in the presence of the program.

[10] Björklund and Moffitt (1987) estimate both the marginal gross gain and the average gross gain from participating in a program. However, they do not present estimates of marginal or average costs.

to the motivation for the simple "before–after" estimator which is still widely used. Its econometric descendent is the fixed effect estimator without a comparison group.

The method assumes that there is access either (i) to longitudinal data on outcomes measured before and after a program for a person who participates in it, or (ii) to repeated cross-section data from the same population where at least one cross-section is from a period prior to the program. To incorporate time into our analysis, we introduce "t" subscripts. Let Y_{1t} be the post-program earnings of a person who participates in the program. When longitudinal data are available, $Y_{0t'}$ is the pre-program outcome of the person. For simplicity, assume that program participation occurs only at time period k, where $t > k > t'$. The "before–after" estimator uses preprogram earnings $Y_{0t'}$ to proxy the no-treatment state in the post-program period. In other words, the underlying identifying assumption is

$$\mathrm{E}(Y_{0t} - Y_{0t'} \mid D = 1) = 0. \tag{4.A.1}$$

If this assumption is valid, the "before–after" estimator is given by

$$(\bar{Y}_{1t} - \bar{Y}_{0t'})_1, \tag{4.1}$$

where the subscript "1" denotes conditioning on $D = 1$, and the bar denotes sample means.

To see how this estimator works, observe that for each individual the gain from the program may be written as

$$Y_{1t} - Y_{0t} = (Y_{1t} - Y_{0t'}) + (Y_{0t'} - Y_{0t}).$$

The second term $(Y_{0t'} - Y_{0t})$ is the approximation error. If this term averages out to zero, we may estimate the impact of participation on those who participate in a program by subtracting participants' mean pre-program earnings from the mean of their post-program earnings. These means also may be defined for different values of participants' characteristics, X.

The before–after estimator does not literally require longitudinal data to identify the means (Heckman and Robb, 1985a,b). As long as the approximation error averages out, repeated cross-sectional data that sample the same population over time, but not necessarily the same persons, are sufficient to construct a before–after estimate. An advantage of this approach is that it only requires information on the participants and their pre-participation histories to evaluate the program.

The major drawback to this estimator is its reliance on the assumption that the approximation errors average out. This assumption requires that among participants, the mean outcome in the no-treatment state is the same in t and t'. Changes in the overall state of the economy between t and t', or changes in the lifecycle position of a cohort of participants, can violate this assumption.

A good example of a case in which assumption (4.A.1) is likely violated is provided in the work of Ashenfelter (1978). Ashenfelter observed that prior to enrollment in a training program, participants experience a decline in their earnings. Later research demonstrates

that Ashenfelter's "dip" is a common feature of the pre-program earnings of participants in government training programs. See Figs. 1–6 which show the dip for a variety of programs in different countries. If this decline in earnings is transitory, and earnings follow a mean-reverting process so that the dip is eventually restored even in the absence of participation in the program, and if period t' falls in the period of transitorily low earnings, then the approximation error will not average out. In this example, the before–after estimator overstates the average effect of training on the trained and attributes mean reversion that would occur under any event to the effect of the program. On the other hand, if the decline is permanent, the before–after estimator is unbiased for the parameter of interest. In this case, any improvement in earnings is properly attributable to the program. Another potential defect of this estimator is that it attributes to the program any trend in earnings due to macro or lifecycle factors.

Two different approaches have been used to solve these problems with the before–after estimators. One controversial method generalizes the before–after estimator by making use of many periods of pre-program data and extrapolating from the period before t' to generate the counterfactual state in period t. It assumes that Y_{0t} and $Y_{0t'}$ can be adjusted to equality using data on the same person, or the same populations of persons, followed over time. As an example, suppose that Y_{0t} is a function of t, or is a function of t-dated variables. If we have

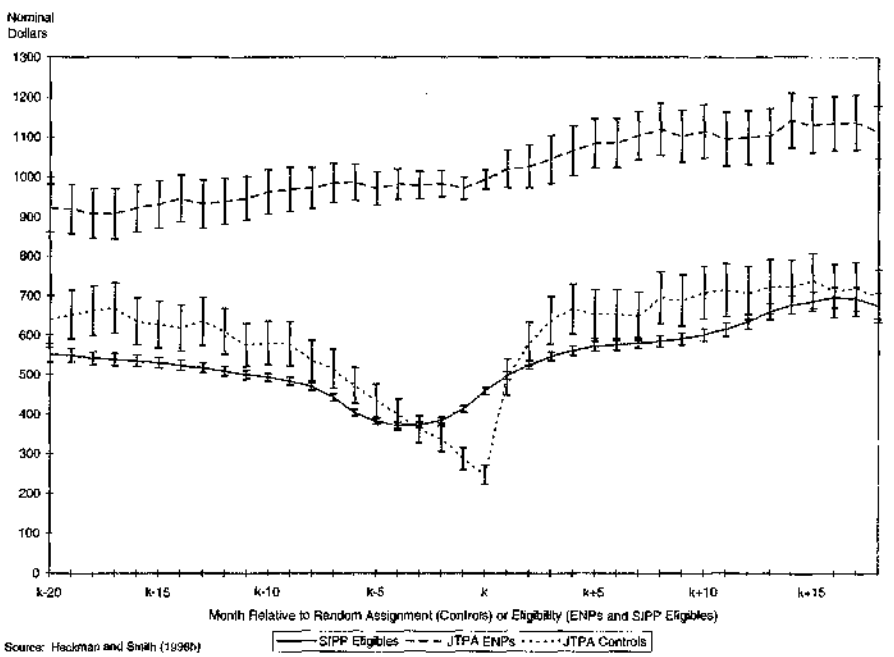

Fig. 1. Mean self-reported monthly earnings: National JTPA Study controls and eligible non-participants (ENPs) and SIPP eligibles (male adults). Source: Heckman and Smith (1999).

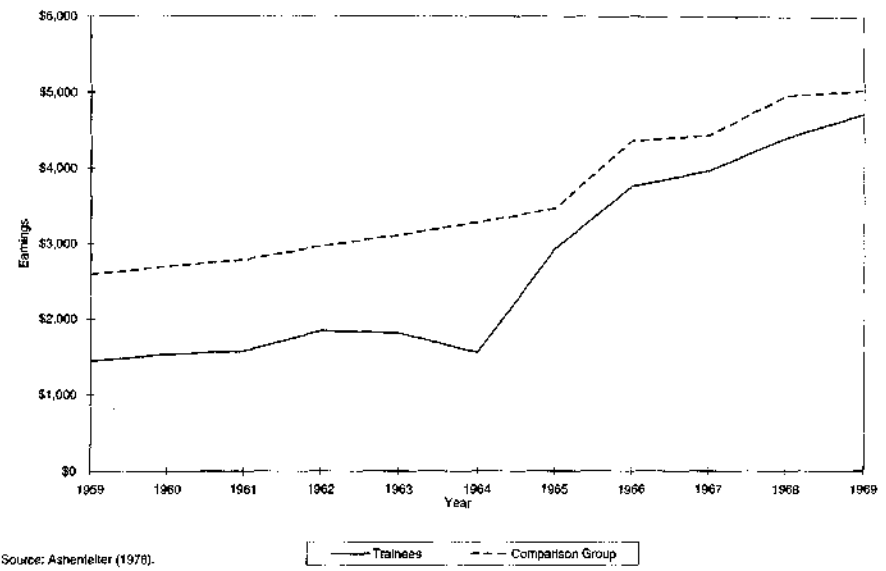

Fig. 2. Mean annual earnings prior, during and subsequent to training for 1964 MDTA classroom trainees and a comparison group (white males).

access to enough data on pre-program outcomes prior to date t' to extrapolate post-program outcomes Y_{0t}, and if there are no errors of extrapolation, or if it is safe to assume that such errors average out to zero across persons in period t, one can replace the missing data or at least averages of the missing data, using extrapolated values. This method is appropriate if population mean outcomes evolve as deterministic functions of time or macroeconomic variables like unemployment. This procedure is discussed further in Section 7.5.[11] The second approach is based on the before–after estimator which we discuss next.

4.2. The difference-in-differences estimator

A more widely used approach to the evaluation problem assumes access either (i) to longitudinal data or (ii) to repeated cross-section data on non-participants in periods t and t'. If the mean change in the no-program outcome measures are the same for participants and non-participants i.e., if the following assumption is valid:

$$\mathsf{E}(Y_{0t} - Y_{0t'} \mid D = 1) = \mathsf{E}(Y_{0t} - Y_{0t'} \mid D = 0), \tag{4.A.2}$$

then the *difference-in-differences* estimator given by

$$(\bar{Y}_{1t} - \bar{Y}_{0t'})_1 - (\bar{Y}_{0t} - \bar{Y}_{0t'})_0, \qquad t > k > t' \tag{4.2}$$

[11] See also Heckman and Robb (1985a, pp. 210–215).

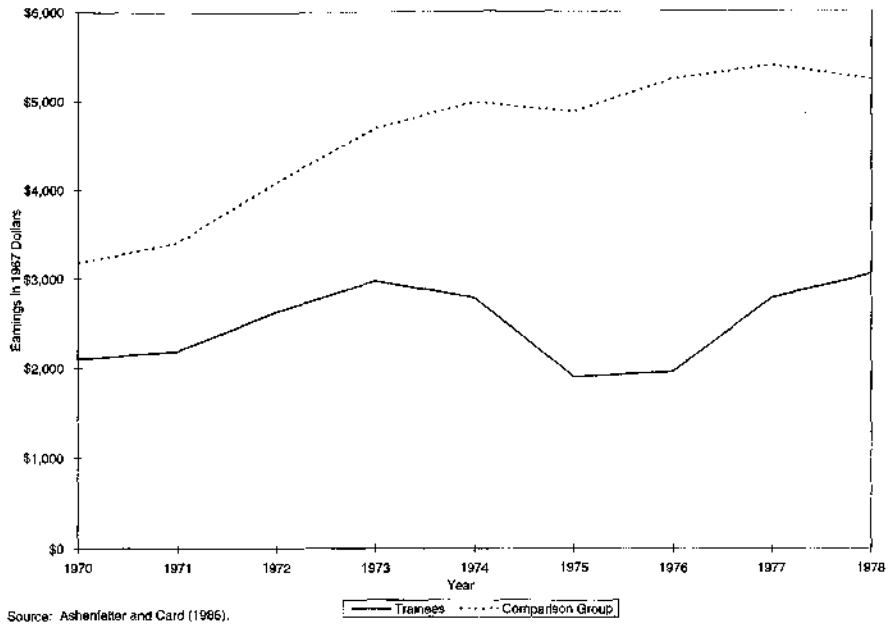

Fig. 3. Mean annual earnings for 1976 CETA trainees and a comparison group (males).

is valid for $E(\Delta_t \mid D = 1) = E(Y_{1t} - Y_{0t} \mid D = 1)$ where $\Delta_t = Y_{1t} - Y_{0t}$ because $E[(\bar{Y}_{1t} - \bar{Y}_{0t'})_1 - (\bar{Y}_{0t} - \bar{Y}_{0t'})_0] = E(\Delta_t \mid D = 1)$.[12] If assumption (4.A.2) is valid, the change in the outcome measure in the comparison group serves to benchmark common year or age effects among participants.

Because we cannot form the change in outcomes between the treated and untreated states, the expression

$$(Y_{1t} - Y_{0t'})_1 - (Y_{0t} - Y_{0t'})_0,$$

cannot be formed for anyone, although we can form one or the other of these terms for everyone. Thus, we cannot use the difference-in-differences estimator to identify the *distribution* of gains without making further assumptions.[13] Like the before–after estimator, we can implement the difference-in-differences estimator for means (4.2) on repeated cross-sections. It is not necessary to sample the same persons in periods t and t' – just persons from the same populations.

[12] The proof is immediate. Make the following decomposition: $(\bar{Y}_{1t} - \bar{Y}_{0t'})_1 = (\bar{Y}_{1t} - \bar{Y}_{0t})_1 + (\bar{Y}_{0t} - \bar{Y}_{0t'})_1$. The claim follows upon taking expectations.

[13] One assumption that identifies the distribution of gains is to assume that $(Y_{1t} - Y_{0t})_1$ is independent of $(Y_{0t} - Y_{0t'})_1$ and that the distribution of $(Y_{1t} - Y_{0t})_1$ is the same as the distribution of $(Y_{0t} - Y_{0t'})_0$. Then the results on deconvolution in Heckman et al. (1997c) can be applied. See their paper for details.

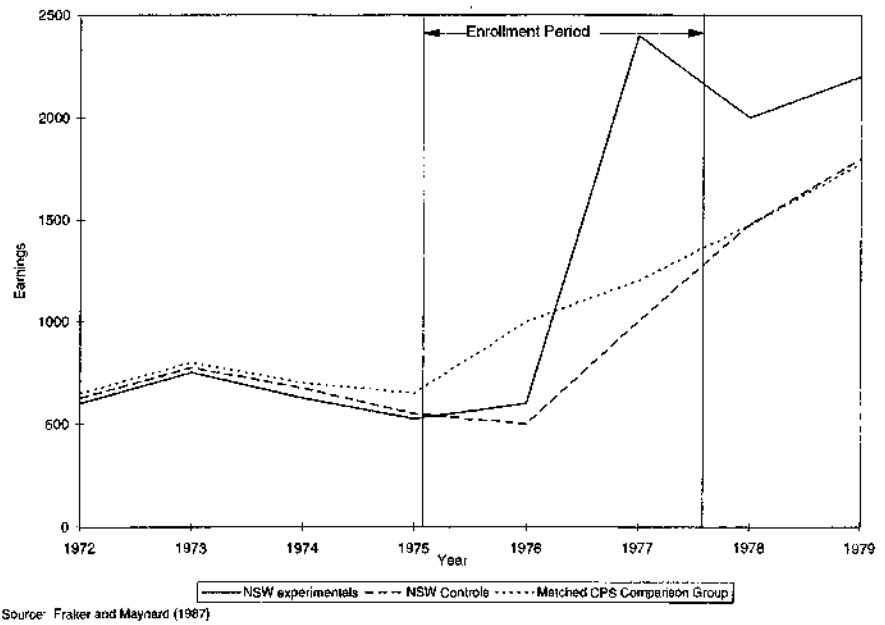

Fig. 4. National supported work (NSW) average annual earnings, treatments, controls and matched CPS comparison group (AFDC recipients).

Ashenfelter's dip provides an example of a case where assumption (4.A.2) is likely to be violated. If Y is earnings, and t' is measured at the time of a transitory earnings dip, and if non-participants do not experience the dip, then (4.A.2) will be violated, because the time path of no-treatment earnings between t' and t will be different between participants and non-participants. In this example, the difference-in-differences estimator overstates the average impact of training on the trainee.

4.3. The cross-section estimator

A third estimator compares mean outcomes of participants and non-participants at time t. This estimator is sometimes called the cross-section estimator. It does not compare the same persons because by hypothesis a person cannot be in both states at the same time. Because of this fact, cross-section estimators cannot estimate the distribution of gains unless additional assumptions are invoked beyond those required to estimate mean impacts.

The key identifying assumption for the cross-section estimator of the mean is that

$$\mathrm{E}(Y_{0t} \mid D = 1) = \mathrm{E}(Y_{0t} \mid D = 0), \tag{4.A.3}$$

i.e., that on average persons who do not participate in the program have the same no-

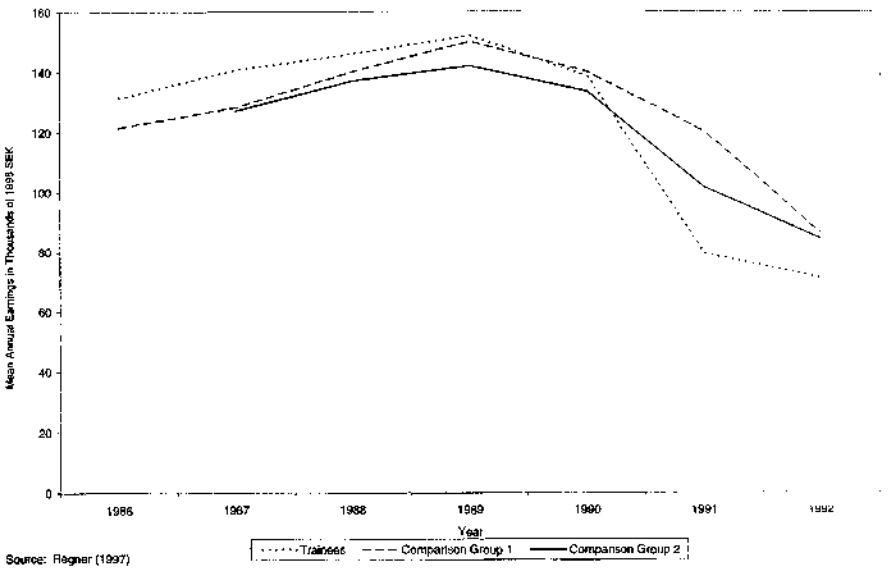

Fig. 5. Earnings of participants in Swedish UI training in 1991 and two comparison groups (adult males aged 26-54).

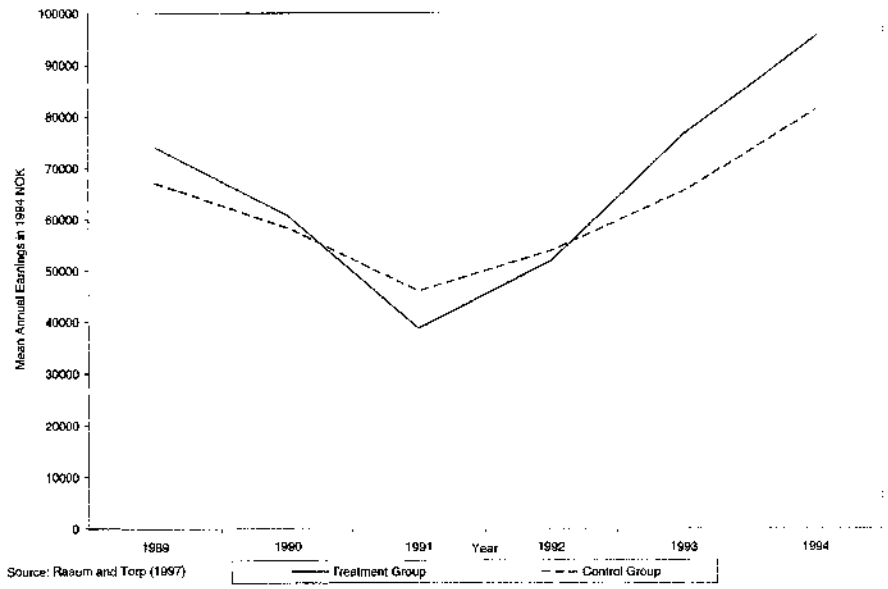

Fig. 6. Earnings of 1991 participants in Norwegian Labor Market Training Program and a randomly assigned control group (all participants).

treatment outcome as those who do participate. If this assumption is valid, then the *cross-section* estimator is given by

$$(\bar{Y}_{1t})_1 - (\bar{Y}_{0t})_0. \tag{4.3}$$

This estimator is valid under assumption (4.A.3) because[14]

$$\mathrm{E}((\bar{Y}_{1t})_1 - (\bar{Y}_{0t})_0) = \mathrm{E}(\Delta_t \mid D = 1).$$

If persons go into the program based on outcome measures in the *post-program* state, then assumption (4.A.3) will be violated. The assumption would be satisfied if participation in the program is unrelated to outcomes in the no-program state *in the post-program period.* Thus, it is possible for Ashenfelter's dip to characterize the data on earnings in the pre-program period, and yet for (4.A.3) to be satisfied. Moreover, as long as the macro economy and aging process operate identically on participants and non-participants, the cross-section estimator is not vulnerable to the problems that plague the before–after estimator.

The cross-section estimator (4.3), the difference-in-differences estimator (4.2), and the before–after estimator (4.1) comprise the trilogy of conventional non-experimental evaluation estimators. All of these estimators can be defined conditional on observable characteristics X. Conditioning on X or additional "instrumental" variables makes it more likely that modified versions of assumptions (4.A.3), (4.A.2), or (4.A.1) will be satisfied but this is not guaranteed. If, for example, the distribution of X characteristics is different between participants ($D = 1$) and non-participants ($D = 0$), conditioning on X may eliminate systematic differences in outcomes between the two groups. Using modern non-parametric procedures, it is possible to exploit each of the identifying conditions to estimate non-parametric versions of all three estimators. On the other hand, if the difference between participants and non-participants is due to unobservables, conditioning may accentuate, and not eliminate, differences between participants and non-participants in the no-program state.[15]

The three estimators exploit three different principles but all are based on making some comparison. The assumptions that justify one method will not, in general, justify any of the other methods. All of the estimators considered in this chapter exploit one of these three principles. They extend the simple mean differences just discussed by making a variety of adjustments to the means. Throughout the rest of the chapter, we organize our discussion of alternative estimators by discussing how they modify the simple mean differences used in the three intuitive estimators to account for non-stationary environments and different values of regressors in the different comparison groups. We first consider social experimentation and how it constructs the counterfactuals used in policy evaluations.

[14] Proof: $(\bar{Y}_{1t})_1 - (\bar{Y}_{0t})_0 = (\bar{Y}_{1t})_1 - (\bar{Y}_{0t})_1 + (\bar{Y}_{0t})_1 - (\bar{Y}_{0t})_0$ and take expectations invoking assumption (4.A.3).

[15] Thus if $\mid \mathrm{E}(Y_0 \mid D = 1) - \mathrm{E}(Y_0 \mid D = 0) \mid = M$, there is no guarantee that $\mid \mathrm{E}(Y_0 \mid D = 1, X) - \mathrm{E}(Y_0 \mid D = 0, X) \mid < M$. For some values of X, the gap could widen.

5. Social experiments

Randomization is one solution to the evaluation problem. Recent years have witnessed increasing use of experimental designs to evaluate North American employment and training programs. This approach has been less common in Europe, though a small number of experiments have been conducted in Britain, Norway and Sweden. When the appropriate qualifications are omitted, the impact estimates from these social experiments are easy for analysts to calculate and for policymakers to understand (see, e.g., Burtless, 1995). As a result of its apparent simplicity, evidence from social experiments has had an important impact on the design of US welfare and training programs.[16] Because of the importance of experimental designs in this literature, in this section we show how they solve the evaluation problem, describe how they have been implemented in practice, and discuss their advantages and limitations.

5.1. How social experiments solve the evaluation problem

An important lesson of this section is that social experiments, like other evaluation methods, provide estimates of the parameters of interest only under certain behavioral and statistical assumptions. To see this, let "*" denote outcomes in the presence of random assignment. Thus, conditional on X for each person we have (Y_1^*, Y_0^*, D^*) in the presence of random assignment and (Y_1, Y_0, D) when the program operates normally without randomization. Let $R = 1$ if a person for whom $D^* = 1$ is randomized into the program and $R = 0$ if the person is randomized out. Thus, $R = 1$ corresponds to the experimental treatment group and R = 0 to the experimental control group.

The essential assumption required to use randomization to solve the evaluation problem for estimating the mean effect of treatment on the treated is that

$$\mathrm{E}(Y_1^* - Y_0^* \mid X, D^* = 1) = \mathrm{E}(Y_1 - Y_0 \mid X, D = 1). \tag{5.A.1}$$

A stronger set of conditions, not strictly required, are

$$\mathrm{E}(Y_1^* \mid X, D^* = 1) = \mathrm{E}(Y_1 \mid X, D = 1) \tag{5.A.2a}$$

and

$$\mathrm{E}(Y_0^* \mid X, D^* = 1) = \mathrm{E}(Y_0 \mid X, D = 1). \tag{5.A.2b}$$

Assumption (5.A.1) states that the means from the treatment and control groups generated by random assignment produce the desired population parameter. With certain exceptions discussed below, this assumption rules out changes in the impact of participation due to the presence of random assignment as well as changes in the process of program participation. The first part of this assumption can in principle be tested by comparing the outcomes of

[16] We discuss this evidence in Section 10.

participants under a regime of randomization with the outcome of participants under the usual regime.

If (5.A.2a) is true, among the population for whom $D = 1$ and $R = 1$ we can identify

$$\mathrm{E}(Y_1 \mid X, D = 1, R = 1) = \mathrm{E}(Y_1 \mid X, D = 1).$$

Under (5.A2a) information sufficient to estimate this mean without bias is routinely produced from data collected on participants in social programs. The new information produced by an experiment comes from those randomized out of the program. Using the experimental control group it is possible to estimate:

$$\mathrm{E}(Y_0 \mid X, D = 1, R = 0) = \mathrm{E}(Y_0 \mid X, D = 1).$$

Simple mean differences identify

$$\mathrm{E}(\Delta \mid X, D = 1) = \mathrm{E}(Y_1 - Y_0 \mid X, D = 1).$$

Within the context of the model of Eq. (3.10), an experiment that satisfies (5.A.1) or (5.A.2a) and (5.A.2b) *does not* make D orthogonal to U. It simply equates the bias in the two groups $R = 1$ and $R = 0$. Thus in the model of Eq. (3.1), under (5.A.2a) and (5.A.2b), $\mathrm{E}(Y \mid X, D = 1, R = 1) = g_1(X) + \mathrm{E}(U_1 \mid X, D = 1)$ and $\mathrm{E}(Y \mid X, D = 1, R = 0) = g_0(X) + \mathrm{E}(U_0 \mid X, D = 1)$.[17]

Rewriting the first conditional mean, we obtain

$$\mathrm{E}(Y \mid X, D = 1, R = 1) = g_1(X) + \mathrm{E}(U_1 - U_0 \mid X, D = 1) + \mathrm{E}(U_0 \mid X, D = 1).$$

Subtracting the second mean from the first eliminates the common selection bias component $\mathrm{E}(U_0 \mid X, D = 1)$ so

$$\mathrm{E}(Y \mid X, D = 1, R = 1) - \mathrm{E}(Y \mid X, D = 1, R = 0) = g_1(X) - g_0(X) + \mathrm{E}(U_1 - U_0 \mid X, D = 1).$$

When the model (3.1) is specialized to one of intercept differences, as in (3.10), this parameter simplifies to α. Notice, that the method of social experiments does *not* set either $\mathrm{E}(U_1 \mid X, D = 1)$ or $\mathrm{E}(U_0 \mid X, D = 1)$ equal to zero. Rather, it balances the selection bias in the treatment and control groups.

Stronger assumptions must be made to identify the distribution of impacts $F(\Delta \mid D = 1)$.[18] Without invoking further assumptions, data from experiments, like data from non-experimental sources, are unable to identify the distribution of impacts because the same person is not observed in both states at the same time (Heckman, 1992; Heckman and Smith, 1993, 1995, 1998a; Heckman et al., 1997c).

If assumption (5.A.1) or assumptions (5.A.2a) and (5.A.2b) fail to hold because the program participation probabilities are affected, so D^* and D are different, then the composition of the participant population differs in the presence of random assignment.

[17] Notice that in this section we allow for the more general model $Y_0 = g_0(X) + U_0$, $Y_1 = g_1(X) + U_1$ where $\mathrm{E}(U_0 \mid X) \neq 0$ and $\mathrm{E}(U_1 \mid X) \neq 0$.

[18] Replace "E" with "F" in (5.A.2a) and (5.A.2b) to obtain one necessary condition.

In two important special cases, experimental data still provide unbiased estimates of the effect of treatment on the treated. First, if the effect of training is the same for everyone, changing the composition of the participants has no effect because the parameter of interest is the same for all possible participant populations (Heckman, 1992). This assumption is sometimes called the common treatment effect assumption and, letting i denote a variable value for individual i, may be formally expressed as

$$Y_{1i} - Y_{0i} = \Delta_i \equiv \Delta, \qquad \text{for all } i. \tag{5.A.3}$$

This assumption is equivalent to setting $U_1 = U_0$ in (3.9). Assumption (5.A.3) can be defined conditionally on observed characteristics, so we may write $\Delta = \Delta(X)$. Notice, however, that in this case, if randomization induces persons with certain X values not to participate in the program, then estimates of $\Delta(X)$ can only be obtained for values of X possessed by persons who participate in the program. In this case (5.A.1) is satisfied but (5.A.2a) and (5.A.2b) are not.

The second special case where experimental data still provide unbiased estimates of the effect of treatment on the treated arises when decisions about training are not affected by the realized gain from participating in the program. This case could arise if potential trainees know $\mathrm{E}(\Delta \mid X)$ but not Δ at the time participation decisions are made. Formally, the second condition is

$$\mathrm{E}(\Delta \mid X, D = 1) = \mathrm{E}(\Delta \mid X), \tag{5.A.4}$$

which is equivalent to condition (3.11) in the model (3.9). If either (5.A.3) or (5.A.4) holds, the simple experimental mean difference estimator is unbiased for $\mathrm{E}(\Delta \mid X,\ D = 1)$.

Randomization improves on the non-experimental cross-section estimator even if there is no selection bias. In an experiment, for all values of X for which $D = 1$, one can identify

$$\mathrm{E}(\Delta \mid X, D = 1) = \mathrm{E}(Y_1 - Y_0 \mid X, D = 1).$$

Using assumption (4.A.3) in an ordinary non-experimental evaluation, there may be values of X such that $\Pr(D = 1 \mid X) = 1$; that is, there may be values of X with no comparison group members. Randomization avoids this difficulty by balancing the distribution of X values in the treatment and control groups (Heckman, 1996). At the same time, however, random assignment conditional on $D = 1$ cannot provide estimates of $\Delta(X)$ for values of X such that $\Pr(D = 1 \mid X) = 0$.

The stage of potential program participation at which randomization is applied – eligibility, application, or acceptance into a program – determines what can be learned from a social experiment. For randomization conditional on acceptance into a program ($D = 1$), we can estimate the effect of treatment on the treated:

$$\mathrm{E}(\Delta \mid X, D = 1) = \mathrm{E}(Y_1 - Y_0 \mid X, D = 1)$$

using simple experimental means. We cannot estimate the effect of randomly selecting a person to go into the program:

$$E(\Delta \mid X) = E(Y_1 - Y_0 \mid X),$$

by using simple experimental means unless one of two conditions prevails. The first condition is just the common effect assumption (5.A.3). This assumption is explicit in the widely used dummy endogenous variable model (Heckman, 1978). The second condition is that embodied in assumption (5.A.4), that participation decisions are independent of the person-specific component of the impact. In both cases, the mean impact of treatment on a randomly selected person is the same as the mean impact of treatment on the treated.

In the general case, it is difficult to estimate the effect of randomly assigning a person with characteristics X to go into a program. This is because persons randomized into a program cannot be compelled to participate in it. In order to secure compliance, it may be necessary to compensate or persuade persons to participate. For example, in many US social experiments, program operators threaten to reduce participants' social assistance benefits, if they refuse to participate in training. Such actions, even if successful, alter the environment in which persons operate and may make it impossible to estimate $E(\Delta \mid X)$ using experimental means. One assumption that guarantees compliance is the existence of a "compensation" or "punishment" level c such that

$$\Pr(D = 1 \mid X, c) = 1 \tag{5.A.5a}$$

and

$$E(\Delta \mid X, c) = E(\Delta \mid X). \tag{5.A.5b}$$

The first part of the assumption guarantees that a person with characteristics X can be "bribed" or "persuaded" to participate in the program. The second part of the assumption guarantees that compensation c does not affect the outcome being evaluated.[19] If c is a monetary payment, it would be optimal from the standpoint of an experimental analyst to find the minimal value of c that satisfies these conditions.

Randomization of eligibility is sometimes proposed as a less disruptive alternative to randomization conditional on $D = 1$. Randomizing eligibility avoids the application and screening costs that are incurred when accepted individuals are randomized out of a program. Because the randomization is performed outside of training centers, it also avoids some of the political costs that have accompanied the use of the experimental method.

Consider a population of persons who are usually eligible for the program. Randomize eligibility within this population. Let $e = 1$ if a person retains eligibility and $e = 0$ if a person becomes ineligible. Assume that eligibility does not disturb the underlying structure of the random variables (Y_0, Y_1, D, X) and that $\Pr(D = 1 \mid X) \neq 0$. Then Heckman (1996) shows that

[19] Observe that the value of c is not necessarily unique.

$$\frac{\mathrm{E}(Y \mid X, e=1) - \mathrm{E}(Y \mid X, e=0)}{\Pr(D=1 \mid X, e=1)} = \mathrm{E}(\Delta \mid X, D=1).$$

Randomization of eligibility produces samples that can be used to identify $\mathrm{E}(\Delta \mid X, D=1)$ and also to recover $\Pr(D=1 \mid X)$. The latter is not recovered from samples which condition on $D=1$ (Heckman, 1992; Moffitt, 1992). Without additional assumptions of the sort previously discussed, randomization on eligibility will not, in general, identify $\mathrm{E}(\Delta \mid X)$.

5.2. *Intention to treat and substitution bias*

The objective of most experimental designs is to estimate the conditional mean impact of training, or $\mathrm{E}(\Delta \mid X, D=1)$. However, in many experiments a significant fraction of the treatment group drops out of the program and does not receive the services being evaluated.[20] In general, in the presence of dropping out $\mathrm{E}(\Delta \mid X, D=1)$ cannot be identified using comparisons of means. Instead, the experimental mean difference estimates the mean effect of the offer of treatment, or what is sometimes called the "intent to treat." For many purposes, this is the policy-relevant parameter. It is informative on how the availability of a program affects participant outcomes. Attrition is a normal feature of an ongoing program.

To obtain an estimate of the impact of training on those who actually receive it, additional assumptions are required beyond (5.A.1) or (5.A.2a) and (5.A.2b). Let T be an indicator for actual receipt of treatment, with $T=1$ for persons actually receiving training, and $T=0$ otherwise. Let T^* be a similarly defined latent variable for control group members indicating whether or not they would have actually received training, had they been in the treatment group. Define

$$\mathrm{E}(\Delta \mid X, D=1, R=1, T=1) = \mathrm{E}(\Delta \mid X, D=1, T=1)$$

as the mean impact of training on those members of the treatment group who actually receive it. This parameter will equal the original parameter of interest $\mathrm{E}(\Delta \mid X, D=1)$ only in the special cases where (5.A.3), the common effect assumption, holds, or where an analog to (5.A.4) holds so that the decision of treatment group members to drop out is independent of $(\Delta - \mathrm{E}(\Delta))$, the person-specific component of their impact.

A consistent estimate of the impact of training on those who actually receive it can be obtained under the assumption that the mean outcome of the treatment group dropouts is the same as that of their analogs in the control group, so that

$$\mathrm{E}(Y \mid X, D=1, R=1, T=0) = \mathrm{E}(Y \mid X, D=1, R=0, T^*=0). \tag{5.A.6}$$

Note that this assumption rules out situations where the treatment group dropouts receive potentially valuable partial treatment. Under (5.A.6),

[20] Using the analysis in the preceding subsection, dropping out by experimental treatment group members could be reduced by compensating them for completing training.

$$\frac{\mathrm{E}(Y \mid X, D=1, R=1) - \mathrm{E}(Y \mid X, D=1, R=0)}{P(T=1 \mid X, D=1, R=1)} \tag{5.1}$$

identifies the mean impact of training on those who receive it.[21] This estimator scales up the experimental mean difference estimate by the fraction of the treatment group receiving training. When all treatment group members receive training, the denominator equals one and the estimator reduces to the simple experimental mean difference. Estimator (5.1) also shows that the simple mean difference estimator provides a downward biased estimate of the mean impact of training on the trained when there are dropouts from the treatment group, because the denominator always lies between zero and one. Heckman et al. (1998f) present methods for estimating distributions of outcomes and for testing the identifying assumptions in the presence of dropping out. They present evidence on the validity of the assumptions that justify (5.1) in the National JTPA Study data.

In an experimental evaluation, the converse problem can also arise for the control group members. In an ideal experiment, no control group members would receive either the experimental treatment or close substitutes to it from other sources. In practice, a significant fraction of controls often receives similar services from other sources. In this situation, the mean earnings of control group members no longer correspond to $\mathrm{E}(Y_0 \mid X, D=1)$ and neither the experimental mean difference estimator nor the adjusted estimator (5.1) identifies the impact of training relative to no training for those who receive it. However, under certain conditions discussed in Section 3, the experimental estimate can be interpreted as the mean incremental effect of the program relative to a world in which it does not exist.

As in the case of treatment group dropouts, identifying the impact of training on the trained in the presence of control group substitution requires additional assumptions beyond (5.A.1) or (5.A.2a) and (5.A.2b). Let $S=1$ denote control group members receiving substitute training from alternative sources and let $S=0$ denote control group members receiving no training and let Y_2 be the outcome conditional on receipt of alternative training. Consider the general case with both treatment group dropping out and control group substitution. In this context, one approach would be to invoke the assumptions required to apply non-experimental techniques as described in Section 7 to the treatment group data to obtain an estimate of the impact of the training being evaluated on those who receive it. Heckman et al. (1998a) employ this and other strategies using data from the National JTPA Study.

Alternatively, two other assumptions allow use of the control group data to estimate the impact of training on the trained. The first assumption is a generalized common effect assumption, where to distinguish individuals we restore subscript i

$$Y_{1i} - Y_{0i} = Y_{2i} - Y_{0i} = \Delta_i \equiv \Delta, \qquad \text{for all } i. \tag{5.A.3$'$}$$

This assumption states that (a) the impact of the program being evaluated is the same as the impact of substitute programs for each person and (b) that all persons respond exactly the

[21] See, e.g., Mallar (1978), Bloom (1984) and Heckman et al. (1998f).

same way to the program (a common effect assumption). The second assumption is a generalized version of (5.A.4), where

$$\mathrm{E}(Y_1 - Y_0 \mid X, D = 1, T = 1, R = 1) = \mathrm{E}(Y_2 - Y_0 \mid X, D = 1, S = 1, R = 0). \qquad (5.\mathrm{A}.4')$$

This assumption states that the mean impact of the training being evaluated received by treatment group members who do not drop out equals the mean impact of substitute training on those control group members who receive it. Both (5.A.3′) and (5.A.4′) are strong assumptions. To be plausible, either would require evidence that the training received by treatment group members was similar in content and duration to that received by control group members. Note that (5.A.3′) implies (5.A.4′). Under either assumption, the ratio

$$\frac{\mathrm{E}(Y \mid X, D = 1, R = 1) - \mathrm{E}(Y \mid X, D = 1, R = 0)}{\Pr(T = 1 \mid X, D = 1, R = 1) - \Pr(S = 1 \mid X, D = 1, R = 0)} \qquad (5.2)$$

identifies the mean impact of training on those who receive it in both the experimental treatment and control groups, provided that the denominator is not zero. The similarity of estimator (5.2) to the instrumental variable estimator defined in Section 7 is not accidental; under assumptions (5.A.3′) or (5.A.4′), random assignment is a valid instrument for training because it is correlated with training receipt but not with any other determinants of the outcome Y. Without one of these assumptions, random assignment is not, in general, a valid instrument (Heckman, 1997; Heckman et al., 1998a). To see this point, consider a model in which individuals know their gain from training, but because the treatment group has access to the program being evaluated, it faces a lower cost of training. In this case, controls are less likely to be trained, but the mean gross impact would be larger among control trainees than among the treatment trainees. Drawing on the analysis of Section 7, this correlation violates the condition required for the IV estimator to identify the parameter of interest.

5.3. Social experiments in practice

In this subsection we discuss how social experiments operate in practice. We present empirical evidence on some of the theoretical issues surrounding social experiments discussed in the preceding subsections and provide a context for the discussion of the experimental evidence on the impact of training in Section 10. To make the discussion concrete, we focus in particular on two of the best known US social experiments: the National Supported Work (NSW) demonstration (Hollister et al., 1984) and the recent National JTPA Study (NJS).[22] We begin with a brief discussion of the implementation of these two experiments.

5.3.1. Two important social experiments

The NSW Demonstration was one of the first employment and training experiments. It tested the effect of 9–18 months of guaranteed work experience in unskilled occupations

[22] See, among others, Doolittle and Traeger (1990), Bloom et al. (1993) and Orr et al. (1994).

on groups of longterm AFDC (welfare) recipients, ex-drug addicts, ex-criminal offenders, and economically disadvantaged youths in 10 sites across the US. These jobs were in a sheltered environment in which productivity standards were gradually raised over time and participants met frequently with program counselors to discuss grievances and performance.

The NSW enrollment process began with a referral, usually by a welfare agency, drug rehabilitation agency, or prisoners' assistance society. Program operators then interviewed potential participants and eliminated any persons that they believed "would be disruptive to their programs" (Hollister et al., 1984, p. 35). Following this screening, a third party randomly assigned one-half of the qualified applicants to the treatment group. The remainder were assigned to the control group and prevented from receiving NSW services. Although the controls could not receive NSW services, program administrators could not prevent them from receiving other training services in their community, such as those offered under another widely available training program with the acronym CETA. Follow-up data on the experimental treatment and control groups were collected via both surveys and administrative earnings records.

In contrast to the NSW, the NJS sought to evaluate the effectiveness of an ongoing training program. From the start, the goal of evaluating an ongoing program without significantly disrupting its operations – and thereby violating assumption (5.A.1) or assumptions (5.A.2a) and (5.A.2b) – posed significant problems. The first of these arose in selecting the training centers at which random assignment would take place. Initially, evaluators planned to use a random sample of the nearly 600 US JTPA training sites. Randomly choosing the evaluation sites would enhance the "external validity" of the experiment – the extent to which its findings can be generalized to the population of JTPA training centers. Yet, it was difficult to persuade local administrators to participate in an evaluation that required them to randomly deny services to eligible applicants. When only four of the randomly selected sites or their alternates agreed to participate, the study was redesigned to include a "diverse" group of 16 centers willing to participate in a random assignment study (see Doolittle and Traeger, 1990; or the summary of their analysis presented in Hotz, 1992). Evaluators had to contact 228 JTPA training centers in order to obtain these sixteen volunteers.[23] The option of forcing centers to participate was rejected because of the importance of securing the cooperation of local administrators in preserving the integrity of random assignment. Such concerns are not without foundation, as the integrity of an experimental training evaluation in Norway was undermined by the behavior of local operators (Torp et al., 1993).

Concerns about disrupting normal program operations and violating (5.A.1) or (5.A.2a)-(5.A.2b) also led to an unusual approach to the evaluation of the specific services provided by JTPA. This program offers a personalized mix of employment and training services including all those listed in Table 1 with the exception of public service employment.

[23] Very large training centers (e.g., Los Angeles) and small, rural centers were excluded from the study design from the outset of the center enrollment process, for administrative and cost reasons, respectively. The final set of 16 training centers received a total of US$1 million in payments to cover the cost of participating in the experiment.

During their enrollment in the program, participants may receive two or more of these services in sequence, where the sequence may depend on the participant's success or failure in those services provided first. As a result of this heterogeneous, fluid structure, it was impossible without changing the character of the program to conduct random assignment conditional on (planned) receipt of particular services or sets of services. Instead, JTPA staff recommended particular services for each potential participant prior to random assignment, and impact estimates were calculated conditional on these recommendations. In particular, the recommendations were grouped into three "treatment streams": the "CT-OS stream" which included persons recommended for classroom training (CT), (and possibly other services), but not on-the-job training (OJT); the "OJT stream" which included persons recommended for OJT (and possibly other services) but not CT; and the "other stream" which included the rest of the admitted applicants, most of whom ended up receiving only job search assistance. Note that this issue did not arise in the NSW, which provided a single service to all of its participants. In the NJS, followup data on earnings, employment and other outcomes were obtained from both surveys and multiple administrative data sources.

5.3.2. The practical importance of dropping out and substitution

The most important problems affecting social experiments are treatment group dropout and control group substitution. These problems are not unique to experiments. Persons drop out of programs whether or not they are experimentally evaluated. There is no evidence that the rate of dropping out increases during an experimental evaluation. Most programs have good substitutes so that the estimated effect of a program as typically estimated is in relation to the full range of activities in which non-participants engage. Experiments exacerbate this problem by creating a pool of persons who attempt to take training who then flock to substitute programs when they are placed in an experimental control group.

Table 3 demonstrates the practical importance of these problems in experimental evaluations by reporting the rates of treatment group dropout and control group substitution from a variety of social experiments. It reveals that the fraction of treatment group members receiving program services is often less than 0.7, and sometimes less than 0.5. Furthermore, the observed characteristics of the treatment group members who drop out often differ from those who remain and receive the program services.[24] In regard to substitution, Table 3 shows that as many as 40% of the controls in some experiments received substitute services elsewhere. In an ideal experiment, all treatments receive the treatment and there is no control group substitution, so that the difference between the fractions of treatments and controls that receive the treatment equals 1.0. In practice, this difference is often well below 1.0.

The extent of both substitution and dropout depends on the characteristics of the treatment being evaluated and the local program environment. In the NSW, where the treat-

[24] For the NSW, see LaLonde (1984); for the NJS see Smith (1992).

ment was relatively unique and of high enough quality to be clearly perceived as valuable by participants, dropout and substitution rates were low enough to approximate the ideal case. In contrast, in the NJS and other evaluations of programs that provide low cost services widely available from other sources, substitution and dropout rates are high.[25] In the NJS, the substitution problem is accentuated by the fact that JTPA relies on outside vendors to provide most of its training. Many of these vendors, such as community colleges, provide the same training to the general public, often with subsidies from other government programs such as Pell Grants. In addition, in order to help in recruiting sites to participate in the NJS, evaluators allowed them to provide control group members with a list of alternative training providers in the community. Of the 16 sites in the NJS, 14 took advantage of this opportunity to alert control group members to substitute training opportunities.

To see the effect of high dropping out and substitution on the interpretation of the experimental evidence, consider Project Independence. The unadjusted experimental impact estimate is $264 over the 2-year followup period, while application of the IV estimator that uses sample moments in place of (5.2) yields an adjusted impact estimate of $1100 (264/0.24). The first estimate indicates the mean impact of the offer of treatment relative to the other employment and training opportunities available in the community. Under assumptions (5.A.3′) or (5.A.4′), the latter estimate indicates the impact of training relative to no training in both the treatment and control groups. Under these assumptions, the high rates of dropping out and substitution suggest that the experimental mean difference estimate is strongly downward biased as an estimate of the impact of treatment on the treated, the primary parameter of policy interest.

A problem unique to experimental evaluations is violation of (5.A.1), or (5.A.2a) and (5.A.2b), which produces what Heckman (1992) and Heckman and Smith (1993, 1995) call "randomization bias." In the NJS, this problem took the form of concerns that expanding the pool of accepted applicants, which was required to keep the number of participants at normal levels while creating a control group, would change the process of selection of persons into the program. Specifically, training centers were concerned that the additional recruits brought in during the experiment would be less motivated and harder to train and therefore benefit less from the program. Concerns about this problem were frequently cited by training centers that declined to participate in the NJS (Doolittle and Traeger, 1990). To partially allay these concerns, random assignment was changed

[25] For the NJS, Table 3 reveals the additional complication that estimates of the rate of training receipt in the treatment and control groups depend on the data source used to make the calculation. In particular, because many treatment group members do not report training that administrative records show they received, dropout rates measured using only the survey data are substantially higher than those that combine the survey and administrative data. At the same time, because administrative data are not available on control group training receipt (other than the very small number of persons who defeated the experimental protocol), using only self-reported data on controls but the combined data for the treatment group will likely overstate the difference in service receipt levels between the two groups.

Table 3
Treatment group dropout and control group substitution in experimental evaluations of active labor market policies (fraction of experimental treatment and control groups receiving services)[a]

Study	Authors/time period	Target group(s)	Fraction of treatments receiving services	Fraction of controls receiving services
1. NSW*	Hollister et al. (1984)	Longterm AFDC women	0.95	0.11
	(9 months after RA)	Ex-addicts	NA	0.03
		17–20 year old HS dropouts	NA	0.04
2. SWIM	Friedlander and Hamilton (1993)	AFDC women: applicants and recipients		
	(Time period not reported)	a. Job search assistance	0.54	0.01
		b. Work experience	0.21	0.01
		c. Classroom training/OJT	0.39	0.21
		d. Any activity	0.69	0.30
		AFDC-U unemployed fathers		
		a. Job search assistance	0.60	0.01
		b. Work experience	0.21	0.01
		c. Classroom training/OJT	0.34	0.22
		d. Any activity	0.70	0.23
3. JOBSTART	Cave et al. (1993)	Youth HS dropouts		
	(12 months after RA)	Classroom training/OJT	0.90	0.26
4. Project Independence	Kemple et al. (1995)	AFDC women: applicants and recipients		
	(24 months after RA)	a. Job search assistance	0.43	0.19
		b. Classroom training/OJT	0.42	0.31
		c. Any activity	0.64	0.40

Table 3 (*continued*)

Study	Authors/time period	Target group(s)	Fraction of treatments receiving services	Fraction of controls receiving services
5. New Chance	Quint et al. (1994) (18 months after RA)	Teenage single mothers		
		Any education services	0.82	0.48
		Any training services	0.26	0.15
		Any education or training	0.87	0.55
6. NJS	Heckman and Smith (1998a,b) (18 months after RA)	Self-reported from Survey Data		
		Adult males	0.38	0.24
		Adult females	0.51	0.33
		Male youth	0.50	0.32
		Female youth	0.58	0.41
		Combined Administrative and Survey Data		
		Adult males	0.74	0.25
		Adult females	0.78	0.34
		Male youth	0.81	0.34
		Female youth	0.81	0.42

[a] Sources: Masters and Maynard (1981, p. 148, Table A.15); Maynard (1980, p. 169, Table A14); Friedlander and Hamilton (1993, p. 22, Table 3.1); Cave et al. (1993, p. 95, Table 4-1); Kemple et al. (1995, p. 58, Table 3.5); Quint et al. (1994, p. 110, Table 4.9); Heckman and Smith (1998a,b) and calculations by the authors. Notes: RA, random assignment; HS, high school. Service receipt includes any employment and training services. The services received by the controls in the NSW study are CETA and WIN jobs. For the longterm AFDC women, this measure also includes regular public sector employment during the period.

from the 1:1 ratio that minimizes the sampling variance of the experimental impact estimator to a 2:1 ratio of treatments to controls.

Although we have no direct evidence on the empirical importance of changes in participation patterns on measured outcomes during the NJS, there is some indirect evidence about the validity of (5.A.1) or (5.A.2a) and (5.A.2b) in this instance. First of all, a number of training centers in the NJS streamlined their intake processes during the experiment - sometimes with the help of an intake consulting firm whose services were subsidized as part of the evaluation. In so doing, they generally reduced the number of visits and other costs paid by potential trainees, thereby including among those randomly assigned less motivated persons than were normally served. Second, some training centers asked for, and received, additional temporary reductions in the random assignment ratio during the course of the experiment when they experienced difficulties recruiting sufficient qualified applicants to keep the program operating at normal levels.

A second problem unique to experiments involves obtaining experimental estimates of the effects of individual components of services provided in sequence as part of a single program. Experimental designs can readily determine how access to a bundle of services affects participants' earnings. More difficult is the question of how participation at each stage influences earnings, when participants can drop out during the sequence. Providing an experimental answer to this question requires randomization at each stage in the sequence.[26] In a program with several stages, this would lead to a proliferation of treatments and either large (and costly) samples or insufficient sample sizes. In practice, such sequential randomization has not been attempted in evaluating job training programs.

A final problem unique to experimental designs is that even under ideal conditions, they are unable to answer many questions of interest besides the narrow impact of "treatment on the treated" parameter. For example, it is not possible in practice to obtain simple experimental estimates of the impact of training on the duration of post-random assignment employment due to post-random assignment selection problems (Ham and LaLonde, 1990). An elaborate analysis of self-selection of the sort sought to be avoided by social experiments is required. As another example, consider estimating the impact of training on wage rates. The problem that arises in this case is that we observe wages only for those employed following random assignment. If the experimental treatment affects employment, then the sample of employed treatments will have different observed and unobserved characteristics than the employed controls. In general, we would expect that the persons without wages will be less skilled. The experimental impact estimate cannot separate out differences between the distributions of observed wages in the treatment and control groups that result from the effect of the program on wage rates from those that result from the effect of the program on selection into employment. Under these

[26] Alternatively, in a program with three stages, program administrators might randomly assign eligible participants to one of several treatment groups, with the first group receiving only stage 1 services, the second receiving stage 1 and stage 2 services and the third receiving services from all three stages. However, a problem may arise with this scheme if participants assigned to the second and third stages of the program at some point decline to participate. In that case, the design described in the text would be more effective.

circumstances, only non-experimental methods such as those discussed in Section 7 can provide an answer to the question of interest.

5.3.3. Additional problems common to all evaluations

There are a number of other problems that arise in both social experiments and non-experimental evaluations. Solving these problems in an experimental setting requires analysts to make the same types of choices (and assumptions) that are required in a non-experimental analysis. An important point of this subsection is that experimental impact estimates are sensitive to these choices in the same way as non-experimental estimates. A related concern is that experimental evaluations should, but often do not, include sensitivity analyses indicating the effect of the choices made on the impact estimates obtained.

The first common evaluation problem arises from imperfect data. Different survey instruments can yield different measures for the same variable for the same person in a given time period (see Smith, 1997a,b, and the citations therein). For example, self-reported measures of earnings or welfare receipt from surveys typically differ from administrative measures covering the same period (LaLonde and Maynard, 1987; Bloom et al., 1993). As we discuss in Section 8, in the case of earnings, data sources commonly used for evaluation research differ in the types of earnings covered, the presence or absence of top-coding and the extent of missing or incorrect values. The evaluator must trade off these factors when choosing which data source to rely on. Whatever the data source used, the analyst must make decisions about how to handle outliers and missing values.

To underscore the point that experimental impacts for the same program can differ due to different choices about data sources and data handling, we compare the impact estimates for the NJS presented in the two official experimental impact reports, Bloom et al. (1993) and Orr et al. (1994).[27] As shown in Table 4, these two reports give substantially different estimates of the impact of JTPA training for the same demographic groups over the same time period. The differences result from different decisions about whom to include in the evaluation sample, how to combine earnings information from surveys and administrative data, how to treat seemingly anomalous reports of overtime earnings in the survey data and so on. Several of the point estimates differ substantially, as do the implications about the relative effectiveness of the three treatment streams for adult women. The estimated 18-month impact for adult women in the "other services" stream triples from the 18-month impact report to the 30-month impact report, making it the service with the largest estimated impact despite the low average cost of the services provided to persons in this stream.

The second problem common to experimental and non-experimental evaluations is sample attrition. Note that sample attrition is not the same as dropping out of the program. Both control and treatment group members can attrit from the sample and treatment group members who drop out of the program will often remain in the data. In the NSW, attrition

[27] A complete discussion of the impact estimates from the NJS appears in Section 10.

Table 4
Variability in experimental impact estimates for adult women in the NJS (mean difference in earnings between the experimental treatment and control groups during the 18 months after random assignment)[a]

Treatment stream	Follow-up report ($)	
	18 month report Bloom et al. (1993)	30 month report Orr et al. (1994)
Recommended for classroom training		
1–6 months	−65	−121
7–18 months	463	312
Sample size	2847	2343
Recommended for on-the-job training		
1–6 months	225	255
7–18 months	518	418
Sample size	2287	2284
Recommended for other services		
1–6 months	171	238
7–18 months	286	879
Sample size	1340	1475

[a] Sources: Bloom et al. (1993, pp. 106, Exhibit 4.12); Orr et al. (1994, pp. 121, 129, 131, Exhibits 5.1, 5.5, and 5.7). Notes: Orr et al. (1994) report the impact per enrollee obtained using the Bloom (1984) estimator rather than the impact per treatment group member. To make the figures in the two columns comparable, we adjusted the impacts per enrollee by the fraction of the treatment group in each recommended service category who enrolled in JTPA. The fraction enrolling among those recommended for classroom training is 0.719, among those recommended for on-the-job training it is 0.532, and among those recommended for other services it is 0.499.

from the evaluation sample by the 18 month followup interview was 10% for the adult women, but more than 30% for the male participants. In the NJS study, sample attrition by the 18 month followup was 12% for the adult women and approximately 20% for the adult males. Such high rates of attrition are common among the disadvantaged due to relatively frequent changes in residence and other difficulties with making followup contacts.

Sample attrition poses a problem for experimental evaluations when it is correlated with individual characteristics or with the impact of treatment conditional on characteristics. In practice, persons with poorer labor market characteristics tend to have higher attrition rates (see, e.g., Brown, 1979). Even if attrition affects both experimental and control groups in the same way, the experiment estimates the mean impact of the program only for those who remain in the sample. Usually, attrition rates are both non-random and larger for controls than for treatments. In this case, the experimental estimate of training is biased because individuals' experimental status, R, is correlated with their likelihood of being in

the sample. In this setting, experimental evaluations become non-experimental evaluations because evaluators must make some assumption to deal with selection bias.

6. Econometric models of outcomes and program participation

The economic approach to program evaluation is based on estimating behavioral relationships that can be applied to evaluate policies not yet implemented. A focus on invariant behavioral relationships is the cornerstone of the econometric approach. Economic relationships provide frameworks within which empirical knowledge can be accumulated across different studies. They offer guidance on the specification of empirical relationships for any given study and the type of data required to estimate a behaviorally-motivated evaluation model. Alternative empirical evaluation strategies can be judged, in part, by the economic justification for them. Estimators that make economically implausible or empirically unjustified assumptions about behavior should receive little support.

The approach to evaluation guided by economic models is in contrast with the case-by-case approach of statistics that at best offers intuitive frameworks for motivating estimators. The emphasis in statistics is on particular estimators and not on the models motivating the estimators. The output of such case-by-case studies often does not cumulate. Since no articulated behavioral theory is used in this approach, it is not helpful in organizing evidence across studies or in suggesting explanatory variables or behaviorally motivated empirical relationships for a given study. It produces estimated parameters that are very difficult to use in answering well-posed evaluation questions.

All economic evaluation models have two ingredients: (a) a model of outcomes and (b) a model of program participation. This section presents several prototypical econometric models. The first was developed by Heckman (1978) to rationalize the evidence in Ashenfelter (1978). The second rationalizes the evidence presented in Heckman and Smith (1999) and Heckman et al. (1998b).

6.1. Uses of economic models

There are several distinct uses of economic models. (1) They suggest lists of explanatory variables that might belong in both outcome and participation equations. (2) They sometimes suggest plausible "exclusion restrictions" - variables that influence participation but do not directly influence outcomes, that can be used to help identify models in the presence of self-selection by participants. (3) They sometimes suggest specific functional forms of estimating equations motivated by a priori theory or by cumulated empirical wisdom.

6.2. Prototypical models of earnings and program participation

To simplify the discussion, and start where the published literature currently stops, assume that persons have only one period in their lives - period k - where they have the chance to take job training. From the beginning of economic life, $t = 1$ up through $t = k$, persons

have one outcome associated with the no-training state "0":

$$Y_{0j}, \qquad j = 1, \ldots, k.$$

After period k, there are two potential outcomes corresponding to the training outcome (denoted "1") and the no-training outcome ("0"):

$$(Y_{0j}, Y_{1j}), \qquad j = k+1, \ldots, T,$$

where T is the end of economic life.

Persons participate in training only if they apply to a program and are accepted into it. Several decision makers may be involved: individuals, family members and bureaucrats. Let $D = 1$ if a person participates in a program; $D = 0$ otherwise. Then the full description of participation and potential outcomes is

$$(D; Y_{0t}, t = 1, \ldots, k; (Y_{0t}, Y_{1t}), t = k+1, \ldots, T). \tag{6.1}$$

As before, observed outcomes after period k can be written as a switching regression model:

$$Y_t = DY_{1t} + (1 - D)Y_{0t}.$$

The most familiar model and the one that is most widely used in the training program evaluation literature assumes that program participation decisions are based on individual choices based on the maximization of the expected present value of earnings. It ignores family and bureaucratic influences on participation decisions.

6.3. *Expected present value of earnings maximization*

In period k, a prospective trainee seeks to measure the expected present value of earnings. Earnings is the outcome of interest. The information available to the agent in period k is I_k. The cost of program participation consists of two components: c (direct costs) and foregone earnings during the training period. Training takes one period to complete. Assume that credit markets are perfect so that agents can lend and borrow freely at interest rate r. The expected present value of earnings maximizing decision rule is to participate in the program ($D = 1$) if

$$\mathrm{E}\left[\sum_{j=1}^{T-k} \frac{Y_{1,k+j}}{(1+r)^j} - c - \sum_{j=0}^{T-k} \frac{Y_{0,k+j}}{(1+r)^j} \mid I_k\right] \geq 0, \tag{6.2}$$

and not to participate in the program ($D = 0$) if this inequality does not hold. In (6.2), the expectations are computed with respect to the information available to the person in period k (I_k). It is important to notice that the expectations in (6.2) are the private expectations of the decision maker. They may or may not conform to the expectations computed against the true ex ante distribution. Note further that I_k may differ among persons in the same environment or may differ among environments. Many variables external to the model

may belong in the information sets of persons. Thus friends, relatives and other channels of information may affect personal expectations.[28]

The following are consequences of this decision rule. (a) Older persons, and persons with higher discount rates, are less likely to take training. (b) Earnings prior to time period k are irrelevant for determining participation in the program except for their value in forecasting future earnings (i.e., except as they enter the person's information set I_k). (c) Only current costs and the discounted gain to earnings determine participation in the program. Persons with lower foregone earnings and lower direct costs of program participation are more likely to go into the program. (d) Any dependence between the realized (measured) income at date t and D is induced by the decision rule. It is the relationship between the expected outcomes at the time decisions are made and the realized outcomes that generate the structure of the bias for any econometric estimator of a model.

This framework underlies much of the empirical work in the literature on evaluating job training programs (see, e.g., Ashenfelter, 1978; Bassi, 1983, 1984; Ashenfelter and Card, 1985). We now consider various specializations of it.

6.3.1. Common treatment effect

As discussed in Section 3, the common treatment effect model is implicitly assumed in much of the literature evaluating job training programs. It assumes that $Y_{1t} - Y_{0t} = \alpha_t$, $t > k$, where α_t is a common constant for everyone. Another version writes α_t as a function of X, $\alpha_t(X)$. We take it as a point of departure for our analysis. The model we first presented was in Heckman (1978). Ashenfelter and Card (1985) and Heckman and Robb (1985a, 1986a) develop it. In this model, the effect of treatment on the treated and the effect of randomly assigning a person to treatment come to the same thing, i.e., $\mathrm{E}(Y_{1t} - Y_{0t} \mid X, D = 1) = \mathrm{E}(Y_{1t} - Y_{0t} \mid X)$ since the difference between the two income streams is the same for all persons with the same X characteristics. Under this model, decision rule (6.2) specializes to the discrete choice model

$$D = 1, \qquad \text{if } \mathrm{E}\left(\sum_{j=1}^{T-k} \frac{\alpha_{k+j}}{(1+r)^j} - c - Y_{0k} \mid I_k\right) \geq 0,$$

$$D = 0, \qquad \text{otherwise.} \tag{6.3}$$

If the α_{k+j} are constant in all periods and T is large ($T \to \infty$) the criterion simplifies to

$$D = 1, \qquad \text{if } \mathrm{E}\left(\frac{\alpha}{r} - c - Y_{0k} \mid I_k\right) \geq 0,$$

$$D = 0, \qquad \text{otherwise.} \tag{6.4}$$

[28] A sharp contrast between a model of perfect certainty and model of uncertainty is that the latter introduces the possibility of incorporating many more "explanatory variables" in the model in addition to the direct objects of the theory.

Even though agents are assumed to be farsighted, and possess the ability to make accurate forecasts, the decision rule is simple. Persons compare current costs (both direct costs c and foregone earnings, Y_{0k}) with expected future rewards

$$\mathrm{E}\left[\left(\sum_{j=1}^{T-k} \frac{\alpha_{k+j}}{(1+r)^j}\right) \mid I_k\right].$$

Future rewards are the same for everyone of the same age and with the same discount rate. Future values of Y_{0t} do not directly determine participation given Y_{0k}. The link between D and Y_{0t}, $t > k$, comes through the dependence with Y_{0k} and any dependence on cost c. If one knew, or could proxy, Y_{0k} and c, one could condition on these variables and eliminate selective differences between participants and non-participants. Since returns are identical across persons, only variation across persons in the direct cost and foregone earnings components determine the variation in the probability of program participation across persons. Assuming that c and Y_{0k} are unobserved by the econometrician, but known to the agent making the decision to go into training,

$$\Pr(D=1) = \Pr\left(\sum_{j=1}^{T-k} \frac{\alpha_{k+j}}{(1+r)^j} > c + Y_{0k}\right).$$

In the case of an infinite-horizon, temporally-constant treatment effect, α, the expression simplifies to

$$\Pr(D=1) = \Pr\left(\frac{\alpha}{r} \geq c + Y_{0k}\right).$$

This simple model is rich enough to be consistent with Ashenfelter's dip. As discussed in Section 4, the "dip" refers to the pattern that the earnings of program participants decline just prior to their participation in the program. If earnings are temporarily low in enrollment period k, and c does not offset Y_{0k}, persons with low earnings in the enrollment period enter the program. Since the return is the same for everyone, it is low opportunity costs or tuition that drive program participation in this model. If the α, c or Y_{0k} depend on observed characteristics, one can condition on those characteristics in constructing the probability of program participation.

This model is an instance of a more general approach to modelling behavior that is used in the economic evaluation literature. Write the net utility of program participation of the decision maker as IN. An individual participates in the program ($D = 1$) if and only if $IN > 0$. Adopting a separable specification, we may write

$$IN = H(X) - V.$$

In terms of the previous example,

$$H(X) = \sum_{j=1}^{T-k} \frac{\alpha_{k+j}}{(1+r)^j}$$

is a constant, and $V = c + Y_{0k}$. The probability that $D = 1$ given X is

$$\Pr(D = 1 \mid X) = \Pr(V < H(X) \mid X). \tag{6.5}$$

If V is stochastically independent of X, we obtain the important special case

$$\Pr(D = 1 \mid X) = \Pr(V < H(X)),$$

which is widely assumed in econometric studies of discrete choice.[29]

If V is normal with mean μ_1 and variance σ_V^2, then

$$\Pr(D = 1 \mid X) = \Pr(V < H(X)) = \Phi\left(\frac{H(X) - \mu_1}{\sigma_V}\right), \tag{6.6}$$

where Φ is the cumulative distribution function of a standard normal random variable. If V is a standardized logit,

$$\Pr(D = 1 \mid X) = \frac{\exp(H(X))}{1 + \exp(H(X))}.$$

Although these functional forms are traditional, they are restrictive and are not required. Conditions for non-parametric identifiability of $\Pr(D = 1 \mid X)$ given different assumptions about the dependence of X and V are presented in Cosslett (1983), and Matzkin (1992). Cosslett (1983), Matzkin (1993) and Ichimura (1993) consider non-parametric estimation of H and the distribution of V. Lewbel (1998) demonstrates how discrete choice models can be identified under much weaker assumptions than independence between X and V. Under certain conditions, information about agent decisions to participate in a training program can be informative about their preferences and the outcomes of a program.

Heckman and Smith (1998a) demonstrate conditions under which knowledge of the self-selection decisions of agents embodied in $\Pr(D = 1 \mid X)$ is informative about the value of Y_1 relative to Y_0. In the Roy model (see, e.g., Heckman and Honoré, 1990), $IN = Y_1 - Y_0 = (\mu_1(X) - \mu_0(X)) + (U_1 - U_0)$. Assuming X is independent of $U_1 - U_0$, from self-selection decisions of persons into a program it is possible to estimate $\mu_1(X) - \mu_0(X)$ up to scale, where the scale is $[\mathrm{Var}(U_1 - U_0)]^{1/2}$. This is a standard result in discrete choice theory. Thus in the Roy model it is possible to recover $\mathrm{E}(Y_1 - Y_0 \mid X)$ up to scale just from knowledge of the choice probability. Under additional assumptions on the support of X, Heckman and Smith (1998a) demonstrate that it is possible to recover the full joint distribution $F(y_0, y_1 \mid X)$ and to answer *all* of the evaluation questions about

[29] Conditions for the existence of a discrete choice random utility representation of a choice process are given in McLennan (1990).

means and distributions posed in Section 3. Under more general self-selection rules, it is still possible to infer the personal valuations of a program from observing selection into the program and attrition from it. The Roy model is the one case where personal evaluations of a program, as revealed by the choice behavior of the agents studied, coincide with the "objective" evaluations based on $Y_1 - Y_0$.

Within the context of a choice-theoretic model, it is of interest to consider the assumptions that justify the three intuitive evaluation estimators introduced in Section 4, starting with the cross-section estimator (4.3) – which is valid if assumption (4.A.3) is correct. Given decision rule (6.3), under what conditions is it plausible to assume that

$$\mathrm{E}(Y_{0t} \mid D = 1) = \mathrm{E}(Y_{0t} \mid D = 0), \qquad t > k \tag{4.A.3}$$

so that cross-section comparisons identify the true program effect? (Recall that in a model with homogeneous treatment impacts, the various mean treatment effects all come to the same thing.) We assume that evaluators do not observe costs nor do they observe Y_{0k} for trainees.

Assumption (4.A.3) would be satisfied in period t if

$$\mathrm{E}\left(Y_{0t} \mid \sum_{j=1}^{T-k} \frac{\alpha_{k+j}}{(1+r)^j} - c - Y_{0k} \geq 0\right) = \mathrm{E}\left(Y_{0t} \mid \sum_{j=1}^{T-k} \frac{\alpha_{k+j}}{(1+r)^j} - c - Y_{0k} < 0\right), \quad t > k.$$

One way this condition can be satisfied is if earnings are distributed independently over time (Y_{0k} independent of Y_{0t}), $t > k$, and direct costs c are independent of Y_{0t}, $t > k$. More generally, only independence in the means with respect to $c + Y_{0k}$ is required.[30] If the dependence in earnings vanishes for earnings measured more than l periods apart (e.g., if earnings are a moving average of order l), then for $t > k + l$, assumption (4.A.3) would be satisfied in such periods.

Considerable evidence indicates that earnings have an autoregressive component (see, e.g., Ashenfelter, 1978; MaCurdy, 1982; Ashenfelter and Card, 1985; Farber and Gibbons, 1994). Then (4.A.3) seems implausible except for special cases.[31] Moreover if stipends (a component of c) are determined in part by current and past income because they are targeted toward low-income workers, then (4.A.3) is unlikely to be satisfied.

Access to better information sometimes makes it more likely that a version of assumption (4.A.3) will be satisfied if it is revised to condition on observables X:

$$\mathrm{E}(Y_{0t} \mid D = 1, X) = \mathrm{E}(Y_{0t} \mid D = 0, X). \tag{4.A.3$'$}$$

In this example, let $X = (c, Y_{0k})$. Then if we observe Y_{0k} for everyone, and can condition on it, and if c is independent of Y_{0t} given Y_{0k}, then

[30] Formally, it is required that $\mathrm{E}(Y_{0t} \mid c + Y_{0k})$ does not depend on c and Y_{0k} for all $t > k$.

[31] Note, however, much of this evidence is for log earnings and not earnings levels.

$$\mathrm{E}(Y_{0t} \mid D = 1, Y_{0k}) = \mathrm{E}\left(Y_{0t} \mid \sum_{j=1}^{T-k} \frac{\alpha_{k+j}}{(1+r)^j} - Y_{0k} \geq c, Y_{0k} \right)$$

$$= \mathrm{E}(Y_{0t} \mid Y_{0k}) = \mathrm{E}(Y_{0t} \mid D = 0, Y_{0k}).$$

Then for common values of Y_{0k}, assumption (4.A.3′) is satisfied for $X = Y_{0k}$.

Ironically, using too much information may make it difficult to satisfy (4.A.3′). To see this, suppose that we observe c and Y_{0k} and $X = (c, Y_{0k})$. Now

$$\mathrm{E}(Y_{0t} \mid D = 1, (c, Y_{0k})) = \mathrm{E}(Y_{0t} \mid c, Y_{0k})$$

and

$$\mathrm{E}(Y_{0t} \mid D = 0, (c, Y_{0k})) = \mathrm{E}(Y_{0t} \mid c, Y_{0k})$$

because c and Y_{0k} perfectly predict D. But (4.A.3′) is *not* satisfied because decision rule (6.3) perfectly partitions the (c, Y_{0k}) space into disjoint sets. There are no common values of $X = (c, Y_{0k})$ such that (4.A.3′) can be satisfied. In this case, the "regression discontinuity design" estimator of Campbell and Stanley (1966) is appropriate. We discuss this estimator in Section 7.4.6.

If we assume that

$$0 < \Pr(D = 1 \mid X) < 1,$$

we rule out the phenomenon of perfect predictability of D given X. This condition guarantees that persons with the same X values have a positive probability of being both participants and non-participants.[32] Ironically, having too much information may be a bad thing. We need some "random" variation that places observationally equivalent people in both states. The existence of this fortuitous randomization lies at the heart of the method of matching.

Next consider assumption (4.A.1). It is satisfied in this example if in a time homogeneous environment, a "fixed effect" or "components of variance structure" characterizes Y_{0t} so that there is an invariant random variable φ such that Y_{0t} can be written as

$$Y_{0t} = \beta_t + \varphi + U_{0t}, \qquad \text{for all } t \tag{6.7}$$

and $\mathrm{E}(U_{0t} \mid \varphi) = 0$ for all t, where the U_{0t} are mutually independent, and c is independent of U_{0t}. If Y_{0t} is earnings, then φ is "permanent income" and the U_{0t} are "transitory deviations" around it. Then using (6.3) for $t > k > t'$, we have

$$\mathrm{E}(Y_{0t} - Y_{0t'} \mid D = 1) = \beta_t - \beta_{t'},$$

since $\mathrm{E}(U_{0t} \mid D = 1) - \mathrm{E}(U_{0t'} \mid D = 1) = 0$.

From the assumption of time homogeneity, $\beta_t = \beta_{t'}$. Thus assumption (4.A.1) is satis-

[32] This is one of two conditions that Rosenbaum and Rubin (1983) call "strong ignorability" and is central to the validity of matching. We discuss these conditions further in Section 7.3.

fied and the before–after estimator identifies α_t. It is clearly not necessary to assume that the U_{0t} are mutually independent, just that

$$\mathrm{E}(U_{0t} - U_{0t'} \mid D = 1) = 0, \tag{6.8}$$

i.e., that the innovation $U_{0t} - U_{0t'}$ is *mean* independent of $U_{0k} + c$. In terms of the economics of the model, it is required that participation does not depend on transitory innovations in earnings in periods t and t'. For decision model (6.3), this condition is satisfied as long as U_{0k} is independent of U_{0t} and $U_{0t'}$, or as long as $U_{0k} + c$ is mean independent of both terms.

If, however, the U_{0t} are serially correlated, then (4.A.1) will generally not be satisfied. Thus if a transitory decline in earnings persists over several time periods (as seems to be true as a consequence of Ashenfelter's dip), so that there is stochastic dependence of $(U_{0t}, U_{0t'})$ with U_{0k}, then it is unlikely that the key identifying assumption is satisfied. One special case where it is satisfied, developed by Heckman (1978) and Heckman and Robb (1985a) and applied by Ashenfelter and Card (1985) and Finifter (1987) among others, is a "symmetric differences" assumption. If t and t' are symmetrically aligned (so that $t = k + l$ and $t' = k - l$) and conditional expectations forward and backward are symmetric, so that

$$\mathrm{E}(U_{0t} \mid c + \beta_k + U_{0k}) = \mathrm{E}(U_{0t'} \mid c + \beta_k + U_{0k}), \tag{6.9}$$

then assumption (4.A.1) is satisfied. This identifying condition motivates the symmetric differences estimator discussed in Section 7.6.

Some evidence of non-stationary wage growth presented by Farber and Gibbons (1994), MaCurdy (1982), Topel and Ward (1992) and others suggests that earnings can be approximated by a "random walk" specification. If

$$Y_{0t} = \beta_t + \eta + \sum_{j=0}^{t} \nu_j, \tag{6.10}$$

where the ν_j are mean zero, mutually independent and identically-distributed random variables independent of η, then (6.8) and (6.9) will not generally be satisfied. Thus even if conditional expectations are linear, both forward and backward, it does not follow that (4.A.1) will hold. Let the variance of η and the variance of ν_j be finite. Assume that $\mathrm{E}(\eta) = 0$. Suppose c is independent of all the ν_j and η, and

$$\mathrm{E}(U_{0t} \mid c + \beta_k + U_{0k}) = \frac{\sigma_\eta^2 + k\sigma_\nu^2}{\sigma_c^2 + \sigma_\eta^2 + k\sigma_\nu^2}(c + U_{0k} - \mathrm{E}(c))$$

and

$$\mathrm{E}(U_{0t'} \mid c + \beta_k + U_{0k}) = \frac{\sigma_\eta^2 + t'\sigma_\nu^2}{\sigma_c^2 + \sigma_\eta^2 + k\sigma_\nu^2}(c + U_{0k} - \mathrm{E}(c)).$$

These two expressions are not equal unless $\sigma_\nu^2 = 0$.

A more general model that is consistent with the evidence reported in the literature writes

$$Y_{0t} = \mu_{0t}(X) + \eta + U_{0t},$$

where

$$U_{0t} = \sum_{j=1}^{k} \rho_{0j} U_{0,t-j} + \sum_{j=1}^{m} m_{0j} \nu_{t-j},$$

where the ν_{t-j} satisfy $E(\nu_{t-j}) = 0$ at all leads and lags, and are uncorrelated with η, and where U_{0t} is an autoregression of order k and moving average of length m. Some authors like MaCurdy (1982) or Gibbons and Farber (1994) allow the coefficients (ρ_{0j}, m_{0j}) to depend on t and do not require that the innovations be identically distributed over time. For the logarithm of white male earnings in the United States, MaCurdy (1982) finds that a model with a permanent component (η), plus one autoregressive coefficient ($k = 1$) and two moving average terms ($m = 2$) describes his data.[33] Gibbons and Farber report similar evidence.

These times series models suggest generalizations of the before–after estimator that exploit the longitudinal structure of earnings processes but work with more general types of differences that align future and past earnings. These are developed at length in Heckman and Robb (1982, 1985a, 1986a), Heckman (1998a) and in Section 7.6.

If there are "time effects," so that $\beta_t \neq \beta_{t'}$, (4.A.1) will not be satisfied. Before–after estimators will confound time effects with program gains. The "difference-in-differences" estimator circumvents this problem for models in which (4.A.1) is satisfied for the unobservables of the model but $\beta_t \neq \beta_{t'}$. Note, however, that in order to apply this assumption it is necessary that time effects be additive in some transformation of the dependent variable and identical across participants and non-participants. If they are not, then (4.A.2) will not be satisfied. For example, if the decision rule for program participation is such that persons with lower lifecycle wage growth paths are admitted into the program, or persons who are more vulnerable to the national economy are trained, then the assumption of common time (or age) effects across participants and non-participants will be inappropriate and the difference-in-differences estimator will not identify true program impacts.

6.3.2. A separable representation

In implementing econometric evaluation strategies, it is common to control for observed characteristics X. Invoking the separability assumption, we write the outcome equation for Y_{0t} as

$$Y_{0t} = g_{0t}(X) + U_{0t},$$

where g_{0t} is a behavioral relationship and U_{0t} has a finite mean conditioning on X. A parallel expression can be written for Y_{1t}:

$$Y_{1t} = g_{1t}(X) + U_{1t}.$$

[33] The estimated value of ρ_{01} is close to 1 so that the model is close is a random walk in levels of log earnings.

The expression for $g_{0t}(X)$ is a structural relationship that may or may not be different from $\mu_{0t}(X)$, the conditional mean. It is a ceteris paribus relationship that informs us of the effect of changes of X on Y_{0t} holding U_{0t} constant. Throughout this chapter we distinguish μ_{1t} from g_{1t} and μ_{0t} from g_{0t}. For the latter, we allow for the possibility that $\mathrm{E}(U_{1t} \mid X) \neq 0$ and $\mathrm{E}(U_{0t} \mid X) \neq 0$. The separability enables us to isolate the effect of self selection, as it operates through the "error term", from the structural outcome equation:

$$\mathrm{E}(Y_{0t} \mid D = 0, X) = g_{0t}(X) + \mathrm{E}(U_{0t} \mid D = 0, X). \tag{6.11a}$$

$$\mathrm{E}(Y_{1t} \mid D = 1, X) = g_{1t}(X) + \mathrm{E}(U_{1t} \mid D = 1, X). \tag{6.11b}$$

The $g_{0t}(X)$ and $g_{1t}(X)$ functions are invariant across different conditioning schemes and decision rules provided that X is available to the analyst. One can borrow knowledge of these functions from other studies collected under different conditioning rules including the conditioning rules that define the samples used in social experiments. Although the conditional mean of the errors differs across studies, the $g_{0t}(X)$ and analogous $g_{1t}(X)$ functions are invariant across studies. If they can be identified, they can be meaningfully compared across studies, unlike the parameter treatment on the treated which, in the case of heterogeneous response to treatment that is acted on by agents, differs across programs with different decision rules and different participant compositions.

A special case of this representation is the basis for an entire literature. Suppose that

(P.1) The random utility representation is valid.

Further, suppose that

(P.2) $(U_{0t}, U_{1t}, V) \perp\!\!\!\perp X$ ($\perp\!\!\!\perp$ denotes stochastic independence)

and finally assume that

(P.3) the distribution of V, $F(V)$, is strictly increasing in V.

Then

$$\mathrm{E}(U_{0t} \mid D = 1, X) = K_{0t}(\Pr(D = 1 \mid X)). \tag{6.12a}$$

and

$$\mathrm{E}(U_{1t} \mid D = 1, X) = K_{1t}(\Pr(D = 1 \mid X)).^{34} \tag{6.12b}$$

[34] The proof is immediate. The proof of (6.12b) follows by similar reasoning. We follow Heckman (1980) and Heckman and Robb (1985a, 1986b). Assume that U_{0t}, V are jointly continuous random variables, with density $f(U_{0t}, V \mid X)$. From (P.2) $f(U_{0t}, V \mid X) = f(U_{0t}, V)$. Thus

$$\mathrm{E}(U_{0t} \mid X, D = 1) = \frac{\int_{-\infty}^{\infty} U_{0t} \int_{-\infty}^{H(X)} f(U_{0t}, V)\mathrm{d}V\, \mathrm{d}U_{0t}}{\int_{\infty}^{H(X)} f(V)\mathrm{d}V}.$$

Now

$$\Pr(D = 1 \mid X) = \int_{\infty}^{H(X)} f(V)\mathrm{d}V.$$

Inverting, we obtain $H(X) = F_V^{-1}(\Pr(D = 1 \mid X))$. Thus

$$\mathrm{E}(U_{0t} \mid X, D = 1) = \frac{\int_{-\infty}^{\infty} U_{0t} \int_{-\infty}^{F_V^{-1}(\Pr(D=1\mid X))} f(U_{0t}, V)\mathrm{d}V\, \mathrm{d}U_{0t}}{\Pr(D = 1 \mid X)} \stackrel{\text{def}}{=} K_{0t}(\Pr(D = 1 \mid X)).$$

The mean error term is a function of P, the probability of participation in the program. This special case receives empirical support in Heckman et al. (1997a, 1998b). It enables analysts to characterize the dependence between U_{0t} and X by the dependence of U_{0t} on $\Pr(D = 1 \mid X)$ which is a scalar function of X. As a practical matter, this greatly reduces the empirical task of estimating selection models. Instead of having to explore all possible dependence relationships between U and X, the analyst can confine attention to the more manageable task of exploring the dependence between U and $\Pr(D = 1 \mid X)$. An investigation of the effect of conditioning on program eligibility rules or self-selection on Y_{0t} comes down to an investigation of the effect of the conditioning on Y_{0t} as it operates through the probability P. It motivates a focus on the determinants of participation in the program in order to understand selection bias and it is the basis for the "control function" estimators developed in Section 7.

If, however, (P.2) is not satisfied, then the separable representation is not valid. Then it is necessary to know more than the probability of participation to characterize $\mathrm{E}(U_{0t} \mid X,\ D = 1)$. In this case it is necessary to characterize both the dependence between U_{0t} and X given $D = 1$ and the probability of participation.

6.3.3. Variable treatment effect

A more general version of the decision rule, given by (6.2), allows (Y_{0t}, Y_{1t}) to be a pair of random variables with no necessary restriction connecting them. In the more general case,

$$\alpha_t = Y_{1t} - Y_{0t}, \qquad t > k$$

is now a random variable. In this case, as previously discussed in Section 3, there is a distinction between the parameter "the mean effect of treatment on the treated" and the "mean effect of randomly assigning a person with characteristics X into the program".

In one important case discussed in Heckman and Robb (1985a), the two parameters have the same ex post mean value even if treatment effect α_t is heterogeneous after conditioning on X. Suppose that α_t is unknown to the agent at the time enrollment decisions are made. The agent forecasts α_t using the information available in his/her information set I_k. $\mathrm{E}(\alpha_t \mid I_k)$ is the private expectation of gain by the agent. If ex post gains of participants with characteristics X are the same as what the ex post gains of non-participants would have been had they participated, then the two parameters are the same. This would arise if both participants and non-participants have the same ex ante expected gains

$$\mathrm{E}(\alpha_t \mid D = 1, I_k) = \mathrm{E}(\alpha_t \mid D = 0, I_k) = \mathrm{E}(\alpha_t \mid I_k),$$

and if

$$\mathrm{E}[\mathrm{E}(\alpha_t \mid I_k) \mid X, D = 1] = \mathrm{E}[\mathrm{E}(\alpha_t \mid I_k) \mid X, D = 0],$$

where the expectations are computed with respect to the observed ex-post distribution of the X. This condition requires that the information in the participant's decision set has the same relationship to X as it has for non-participants. The interior expectations in the

preceding expression are subjective. The exterior expectations in the expression are computed with respect to distributions of objectively observed characteristics. The condition for the two parameters to be the same is

$$\mathrm{E}[\mathrm{E}(\alpha_t \mid I_k, D = 1) \mid X, D = 1] = \mathrm{E}[\mathrm{E}(\alpha_t \mid I_k, D = 0) \mid X, D = 0].$$

As long as the ex-post objective expectation of the subjective expectations is the same, the two parameters $\mathrm{E}(\alpha_t \mid X,\ D = 1)$ and $\mathrm{E}(\alpha_t \mid X)$ are the same. This condition would be satisfied if, for example, all agents, irrespective of their X values, place themselves at the mean of the objective distribution, i.e.,

$$\mathrm{E}(\alpha_t \mid I_k, D = 1) = \mathrm{E}(\alpha_t \mid I_k, D = 0) = \bar{\alpha}_t$$

(see, e.g., Heckman and Robb, 1985a). Differences across persons in program participation are generated by factors other than potential outcomes. In this case, the ex-post surprise,

$$(\alpha_t - \bar{\alpha}_t)$$

does not depend on X or D in the sense that

$$\mathrm{E}(\alpha_t - \bar{\alpha}_t \mid X, D = 1) = 0.$$

So

$$\mathrm{E}(Y_{1t} - Y_{0t} \mid X, D = 1) = \bar{\alpha}_t.$$

This discussion demonstrates the importance of understanding the decision rule and its relationship to measured outcomes in formulating an evaluation model. If agents do not make their decisions based on the unobserved components of gains from the program or on variables statistically related to those components, the analysis for the common coefficient model presented in section (a) remains valid even if there is variability in $U_{1t} - U_{0t}$. If agents anticipate the gains, and base decisions on them, at least in part, then a different analysis is required.

The conditions for the absence of bias for one parameter are different from the conditions for the absence of bias for another parameter. The difference between the "random assignment" parameter $\mathrm{E}(Y_{1t} - Y_{0t} \mid X)$ and the "treatment on the treated" parameter is the gain in the unobservables going from one state to the next:

$$\mathrm{E}(U_{1t} - U_{0t} \mid X, D = 1) = \mathrm{E}(\Delta_t \mid X, D = 1) - \mathrm{E}(\Delta_t \mid X).$$

The only way to avoid bias for *both* mean parameters is if $\mathrm{E}(U_{1t} - U_{0t} \mid X,\ D = 1) = 0$.

Unlike the other estimators, the before–after estimators are non-robust to time effects that are common across participants and non-participants. The difference-in-differences estimators and the cross-section estimators are unbiased under different conditions. The cross-section estimators for the period t common effect and the "treatment on the treated" variable-effect version of the model require that mean unobservables in the no-program state be the same for participants and non-participants. The difference-in-differences

estimator requires a *balance of the bias* in the *change* in the unobservables from period t' to period t. If the cross-section conditions for the absence of bias are satisfied for all t, then the assumption justifying the difference-in-differences estimator is satisfied.

However, the converse is not true. Even if the conditions for the absence of bias in the difference-in-differences estimator are satisfied, the conditions for absence of bias for the cross-section estimator are not necessarily satisfied. Moreover, failure of the difference-in-differences condition for the absence of bias does not imply failure of the condition for absence of bias for the cross-section estimator. Ashenfelter's dip provides an empirically relevant example of this point. If t' is measured during the period of the dip, but the dip is mean-reverting in post-program periods, then the condition for the absence of cross-section bias could be satisfied because post-program, there could be no selective differences among participants.

6.3.4. Imperfect credit markets

How robust is the analysis of Sections 6.2 and 6.3, and in particular the conditions for bias, to alternative specifications of decision rules and the economic environments in which individuals operate? To answer this question, we first reexamine the decision rule after dropping our assumption of perfect credit markets. There are many ways to model imperfect credit markets. The most extreme approach assumes that persons consume their earnings each period. This changes the decision rule (6.2) and produces a new interpretation for the conditions for absence of bias. Let G denote a time-separable strictly concave utility function and let β be a subjective discount rate. Suppose that persons have exogenous income flow η_t per period. Expected utility maximization given information I_k produces the following program participation rule:

$$D = \begin{cases} 1 \text{ if } \mathrm{E}\left[\sum_{j=1}^{T-k} \beta^j G(Y_{1,k+j} + \eta_{k+j}) - G(Y_{0,k+j} + \eta_{k+j}) + G(\eta_k - c_k) - G(Y_{0k} + \eta_k) \mid I_k\right] \geq 0; \\ 0 \quad \text{otherwise.} \end{cases} \tag{6.13}$$

As in the previous cases, earnings prior to time period k are only relevant for forecasting future earnings (i.e., as elements of I_k). However, the decision rule (6.2) is fundamentally altered in this case. Future earnings in both states determine participation in a different way. Common components of earnings in the two states do not difference out unless G is a linear function.[35]

Consider the permanent-transitory model of Eq. (6.7). That model is favorable to the application of longitudinal before–after estimators. Suppose that the U_{0t} are independent and identically distributed, and there is a common-effect model. Condition (6.8) is not

[35] Due to the non-linearity of G, there are wealth effects in the decision to take training.

satisfied in a perfect foresight environment when there are credit constraints, or in an environment in which the U_{0t} can be partially forecast,[36] because for $t > k > t'$

$$E(U_{0t} \mid X, D = 1) \neq 0$$

even though

$$E(U_{0t'} \mid X, D = 1) = 0$$

so

$$E(U_{0t} - U_{0t'} \mid X, D = 1) \neq 0.$$

The before–after estimator is now biased. So is the difference-in-differences estimator. If, however, the U_{0t} are not known, and cannot be partially forecast, then condition (6.8) is valid, so both the before–after and difference-in-differences estimators are unbiased.

Even in a common effect model, with Y_{0t} (or U_{0t}) independently and identically distributed, the cross-section estimator is biased for period $t > k$ in an environment of perfect certainty with credit constraints because D depends on Y_{0t} through decision rule (6.13). On the other hand, if Y_{0t} is not forecastable with respect to the information in I_k, the cross-section estimator is unbiased.

The analysis in this subsection and the previous subsections has major implications for a certain style of evaluation research. Understanding the stochastic model of the outcome process is not enough. It is also necessary to know how the decision-makers process the information, and make decisions about program participation.

6.3.5. Training as a form of job search

Heckman and Smith (1999) find that among persons eligible for the JTPA program, the unemployed are much more likely to enter the program than are other eligible persons. Persons are defined to be unemployed if they are not working but report themselves as actively seeking work. The relationship uncovered by Heckman and Smith is not due to eligibility requirements. In the United States, unemployment is not a precondition for participation in the program.

Several previous studies suggest that Ashenfelter's dip results from changes in labor force status, instead of from declines in wages or hours among those who work. Using even a crude measure of employment rates, namely whether a person was employed at all during a calendar year, Card and Sullivan (1988) observed that US CETA training parti-

[36] "Partially forecastable" means that some component of U_{0t} resides in the information set I_k. That is, letting $f(y \mid x)$ be the density of Y given X, $f(U_{0t} \mid I_k) \neq f(U_{0t})$ so that I_k predicts U_{0t} in this sense. One could define "moment forecastability" using conditional expectations of certain moments of function "φ". If $E(\varphi(U_{0t}) \mid I_k) \neq E(\varphi(U_{0t}))$, then $\varphi(U_{0t})$ is partially moment forecastable using the information in I_k. More formally, a random variable is fully-forecastable if the σ-algebra generating U_{0t} is contained in the σ-algebra of I_k. It is partially forecastable if the complement of the projection of the σ-algebra of U_{0t} onto the σ-algebra of I_k is not the empty set. It is fully unforecastable if the projection of the σ-algebra of U_{0t} onto the σ-algebra of I_k is the empty set.

cipants' employment rates declined prior to entering training.[37] Their evidence suggests that changes in labor force dynamics instead of changes in earnings may be a more precise way to characterize participation in training.

Heckman and Smith (1999) show that whether or not a person is employed, unemployed (not employed and looking for work), or out of the labor force is a powerful predictor of participation in training programs. Moreover, they find that recent changes in labor force status are important determinants of participation for all demographic groups. In particular, eligible persons who have just become unemployed, either through job loss or through re-entry into the labor force, have the highest probabilities of participation. For women, divorce, another form of job termination, is a predictor of who goes into training. Among those who either are employed or out of the labor force, persons who have recently entered these states have much higher program participation probabilities than persons in those states for some time. Their evidence is formalized by the model presented in this section.

The previous models that we have considered are formulated in terms of *levels* of costs and earnings. When opportunity costs are low, or tuition costs are low, persons are more likely to enter training. The model presented here recognizes that *changes* in labor force states account for participation in training. Low earnings levels are a subsidiary predictor of program participation that are overshadowed in empirical importance by unemployment dynamics in the analyses of Heckman and Smith (1999).

Persons with zero earnings differ substantially in their participation probabilities depending on their recent labor force status histories. Yet, in models based on pre-training earnings dynamics, such as the one presented in Section 6.3, such persons are assumed to have the same behavior irrespective of their labor market histories.

The importance of labor force status histories also is not surprising given that many employment and training services, such as job search assistance, on-the-job training at private firms, and direct placement are all designed to lead to immediate employment. By providing these services, these programs function as a form of job search for many participants. Recognizing this role of active labor market policies is an important development in recent research. It indicates that in many cases, participation in active labor market programs should not be modeled as if it were like a schooling decision, such as we have modeled it in the preceding sections.

In this section, we summarize the evidence on the determinants of participation in the program and construct a simple economic model in which job search makes two contributions to labor market prospects: (a) it increases the rate of arrival of job offers and (b) it improves the distribution of wages in the sense of giving agents a stochastically dominant wage distribution compared to the one they face without search. Training is one form of unemployment that facilitates job search. Different training options will produce different job prospects characterized by different wage and layoff distributions. Searchers might participate in programs that subsidize the rate of arrival of job offers (JSA as described in

[37] Ham and LaLonde (1990) report the same result using semi-monthly employment rates for adult women participating in NSW.

Section 2), or that improve the distribution from which wage offers are drawn (i.e., basic educational and training investments).

Instead of motivating participation in training with a standard human capital model, we motivate participation as a form of search among options. Because JSA constitutes a large component of active labor market policy, it is of interest to see how the decision rule is altered if enhanced job search rather than human capital accumulation is the main factor motivating individuals' participation in these programs.

Our model is based on the idea that in program j, wage offers arrive from a distribution F_j at rate λ_j. Persons pay c_j to sample from F_j. (The costs can be negative). Assume that the arrival times are statistically independent of the wage offers and that arrival times and wage offers from one search option are independent of the wages and arrival times of other search options. At any point in time, persons pick the search option with the highest expected return. To simplify the analysis, suppose that all distributions are time invariant and denote by N the value of non-market time. Persons can select among any of J options, denoted by j. Associated with each option is a rate at which jobs appear, λ_j. Let the discount rate be r. These parameters may vary among persons but for simplicity we assume that they are constant for the same person over time. This heterogeneity among persons produces differences among choices in training options, and differences in the decision to undertake training.

In the unemployed state, a person receives a non-market benefit, N. The choice between search from any of the training and job search options can be written in "Gittens Index" form (see, e.g., Berry and Fristedt, 1985). Under our assumptions, being in the non-market state has constant per-period value N irrespective of the search option selected. Letting V_{je} be the value of employment arising from search option j, the value of being unemployed under training option j is

$$V_{ju} = N - c_j + \frac{\lambda_j}{1+r}\mathrm{E}_j \max[V_{je}; V_{ju}] + \frac{(1-\lambda_j)}{1+r}V_{ju}. \tag{6.14a}$$

The first term, $(N - c_j)$, is the value of non-market time minus the j-specific cost of search. The second term is the discounted product of the probability that an offer arrives next period if the jth option is used, and the expected value of the maximum of the two options: work (valued at V_{je}) or unemployment (V_{ju}). The third term is the probability that the person will continue to search times the value of doing so. In a stationary environment, if it is optimal to search from j today, it is optimal to do so tomorrow.

Let σ_{je} be the exogenous rate at which jobs disappear. For a job holder, the value of employment is V_{je}:

$$V_{je} = Y_j + \frac{(1-\sigma_{je})}{1+r}V_{je} + \frac{\sigma_{je}}{1+r}\mathrm{E}_j[\max(V_N, V_{ju})]. \tag{6.14b}$$

V_{ju} is the value of optimal job search under j. The expression consists of the current flow of earnings (Y_j) plus the discounted ($1/1 + r$) expected value of employment (V_{je}) times the probability that the job is retained ($1 - \sigma_{je}$). The third term arises from the possibility that

a person loses his/her job (this happens with probability (σ_{je})) times the expected value of the maximum of the search and non-market value options (V_N).

To simplify this expression, assume that $V_{ju} > V_N$. If this is not so, the person would never search under any training option under any event. In this case, V_{je} simplifies to

$$V_{je} = Y_j + \frac{(1 - \sigma_{je})}{1 + r} V_{je} + \frac{\sigma_{je}}{1 + r} V_{ju}$$

so

$$V_{je} = \frac{\sigma_{je}}{r + \sigma_{je}} V_{ju} + \frac{(1 + r)Y_j}{r + \sigma_{je}}. \tag{6.14c}$$

Substituting (6.14c) into (6.14a), we obtain, after some rearrangement,

$$V_{ju} = \frac{(1 + r)(N - c_j) + \lambda_j \mathrm{E}_j(V_{je} \mid V_{je} > V_{ju}) \Pr(Y_j > V_{ju}(r/(1 + r)))}{r + \lambda_j \Pr(Y_j > V_{ju}(r/(1 + r)))}.$$

In deriving this expression, we assume that the environment is stationary so that the optimal policy at time t is also the optimal policy at t' provided that the state variables are the same in each period.

The optimal search strategy is

$$\hat{j} = \underset{j}{\operatorname{argmax}} \{V_{ju}\}$$

provided that $V_{ju} > V_N$ for at least one j. The lower c_j and the higher λ_j, the more attractive is option j. The larger the F_j – in the sense that j stochastically dominates j' ($F_j(x) < F_{j'}(x)$), so more of the mass of F_j is the upper portion of the distribution – the more attractive is option j. Given the search options available to individuals, enrollment in a job training program may be the most effective option.

The probability that training from option j lasts $T_j = t_j$ periods or more is

$$\Pr(T_j \geq t_j) = [1 - \lambda_j(1 - F_j(V_{ju}(r/(1 + r))))]^{t_j},$$

where $1 - \lambda_j(1 - F_j(V_{ju}(r/(1 + r))))$ is the sum of the probability of receiving no offer ($1 - \lambda_j$) plus the probability of receiving an offer that is not acceptable ($\lambda_j F_j(V_{ju}(r/(1 + r)))$). This model is non-linear in the basic parameters. Because of this non-linearity, many estimators relying on additive separability of the unobservables, such as difference-in-differences or the fixed effect schemes for eliminating unobservables, are ineffective evaluation estimators.

This simple model summarizes the available empirical evidence on job training programs. (a) It rationalizes variability in the length of time persons with identical characteristics spend in training. Persons receive different wage offers at different times and leave the program to accept the wage offers at different dates. (b) It captures the notion that training programs might facilitate the rate of job arrivals – the λ_j (this is an essential function of "job search assistance" programs) or they might produce skills – by improving

the F_j – or both. (c) It accounts for why there might be recidivism back into training programs. As jobs are terminated (at rate σ_{je}), persons re-enter the program to search for a replacement job. Recidivism is an important feature of major job training programs. Trott and Baj (1993) estimate that as many as 20% of all JTPA program participants in Northern Illinois have been in the program at least twice with the modal number being three. This has important implications for the contamination bias problem that we discuss in Section 7.7.

A less attractive feature of the model is that persons do not switch search strategies. This is a consequence of the assumed stationarity of the environment and the assumption that agents know both arrival rates and wage offer distributions. Relaxing the stationarity assumption produces switching among strategies which seems to be consistent with the evidence. A more general – but less analytically tractable model – allows for learning about wage offer distributions as in Weitzman (1979). In such a model, persons may switch strategies as they learn about the arrival rates or the wage offers obtained under a given strategy. The learning can take place within each type of program and may also entail word of mouth learning from fellow trainees taking the option.

Weitzman's model captures this idea in a very simple way and falls within the Gitten's index framework. The basic idea is as follows. Persons have J search options. They pick the option with the highest value and take a draw from it. They accept the draw if the value of the realized draw is better than the expected value of the best remaining option. Otherwise they try out the latter option. If the draws from the J options are independently distributed, a Gittens-index strategy describes this policy. In this framework, unemployed persons may try a variety of options – including job training – before they take a job, or drop out of the labor force.

One could also extend this model to allow the value of non-market time, N, to become stochastic. If N fluctuates, persons would enter or exit the labor force depending on the value of N. Adding this feature captures the employment dynamics of trainees described by Card and Sullivan (1988).

In this more general model, shocks to the value of leisure or termination of previous jobs make persons contemplate taking training. Whether or not they do so depends on the value of training compared to the value of other strategies for finding jobs. Allowing for these considerations produces a model broadly consistent with the evidence presented in Heckman and Smith (1999) that persons enter training as a consequence of displacement from both the market and non-market sector.

The full details of this model remain to be developed. We suggest that future analyses of program participation be based on this empirically more concordant model. For the rest of this chapter, however, we take decision rule (6.2) as canonical in order to motivate and justify the choice of alternative econometric estimators. We urge our readers to modify our analysis to incorporate the lessons from the framework of labor force dynamics sketched here.

6.4. The role of program eligibility rules in determining participation

Several institutional features of most training programs suggest that the participation rule is more complex than that characterized by the simple model presented above in Section 6.3. For example, eligibility for training is often based on a set of objective criteria, such as current or past earnings being below some threshold. In this instance, individuals can take training at time k only if they have had low earnings, regardless of its potential benefit to them. For example, enrollees satisfy

$$\alpha/r - Y_{ik} - c_i > 0 \tag{6.15}$$

and the eligibility rule $Y_{i,k-1} < K$ where K is a cutoff level. More general eligibility rules can be analyzed in the same framework.

The universality of Ashenfelter's dip in pre-program earnings among program participants occurs despite the substantial variation in eligibility rules among training programs. This suggests that earnings or employment dynamics drive the participation process and that Ashenfelter's dip is not an artifact of eligibility rules. Few major training programs in the United States have required earnings declines to qualify for program eligibility. Certain CETA programs in the late 1970s required participants to be unemployed during the period just prior to enrollment, while NSW required participants to be unemployed at the date of enrollment. MDTA contained no eligibility requirements, but restricted training stipends to persons who were unemployed or "underemployed."[38] For the JTPA program, eligibility has been confined to the economically disadvantaged (defined by low family income over the past 6 months, participation in a cash welfare program or Food Stamps or being a foster child or disabled). There is also a 10% "audit window" of eligibility for persons facing other unspecified "barriers to employment."

It is possible that Ashenfelter's dip results simply from a mechanical operation of program eligibility rules that condition on recent earnings. Such rules select individuals with particular types of earnings patterns into the eligible population. To illustrate this point, consider the monthly earnings of adult males who were eligible for JTPA in a given month from the 1986 panel of the US Survey of Income and Program Participation (SIPP). For most people, eligibility is determined by family earnings over the past 6 months. The mean monthly earnings of adult males appear in Fig. 1 aligned relative to month k, the month when eligibility is measured. The figure reveals a dip in the mean earnings of adult

[38] Eligibility for CETA varied by subprogram. CETA's controversial Public Sector Employment (PSE) program required participants to have experienced a minimum number of days of unemployment or "underemployment" just prior to enrollment. In general, persons became eligible for other CETA programs by having a low income or limited ability in English. Considerable discretion was left to the states and training centers to determine who enrolled in the program. By contrast, the NSW eligibility requirements were quite specific. Adult women had to be on AFDC at the time of enrollment, have received AFDC for 30 of the last 36 months, and have a youngest child age 6 years or older. Youth in the NSW had to be age 17–20 years with no high school diploma or equivalency degree and have not been in school in the past 6 months. In addition, 50% of youth participants had to have had some contact with the criminal justice system (Hollister et al., 1984).

male *eligibles* centered in the middle of the six month window over which family income is measured when determining JTPA eligibility.

Fig. 1 also displays the mean earnings of adult males in the experimental control group from the NJS.[39] The earnings dip for the controls, who applied and were admitted to the program, is larger than for the sample of JTPA eligibles from the SIPP. Moreover, this dip reaches its minimum during month k rather than 3 or 4 months before as would be indicated by the operation of eligibility rules. The substantial difference between the mean earnings patterns of JTPA participants and eligibles implies that Ashenfelter's dip does not result from the mechanical operation of program eligibility rules.[40]

6.5. Administrative discretion and the efficiency and equity of training provision

Training participation also often depends on discretionary choices made by program operators. Recent research focuses on how program operators allocate training services among groups and on how administrative performance standards affect the allocation of these services. The main question that arises in these studies is the potential tradeoff between equity and efficiency, and the potential conflict between social objectives and program operators' incentives. An efficiency criterion that seeks to maximize the social return to public training investments, regardless of the implications for income distribution, implies focusing training resources on those groups for whom the impact is largest (per dollar spent). In contrast, equity and redistributive criteria dictate focusing training resources on groups who are most in "need" of services.

These goals of efficiency and equity are written into the US Job Training Partnership Act.[41] Whether or not these twin goals conflict with each other depends on the empirical relationship between initial skill levels and the impact of training. As we discuss below in Section 10, the impact of training appears to vary on the basis of observable characteristics, such as sex, age, race and what practitioners call "barriers to employment" -- low schooling, lack of employment experience and so on. These twin goals would be in conflict if the largest social returns resulted from training the most job-ready applicants.

In recent years, especially in the United States, policymakers have used administrative performance standards to assess the success of program operators in different training sites. Under JTPA, these standards are based primarily on average employment rates and average wage rates of trainees shortly after they leave training. The target levels for each site are adjusted based on a regression model that attempts to hold constant features of the

[39] Such data were collected at four of the 16 training centers that participated in the study.

[40] Devine and Heckman (1996) present certain non-stationary family income processes that can generate Ashenfelter's dip from the application of JTPA eligibility rules. However, in their empirical work they find a dip centered at $k-3$ or $k-4$ for adult men and adult women, but no dip for male and female youth.

[41] A related issue involves differences in the types of services provided to different groups conditional on participation in a program. The US General Accounting Office (1991) finds such differences alarming in the JTPA program. Smith (1992) argues that they result from differences across groups in readiness for immediate employment and in the availability of income support during classroom training.

environment over which the local training site has no control, such as racial composition.[42] Sites whose performance exceeds these standards may be rewarded with additional funding; those that fall below may be sanctioned. The use of such performance standards, instead of measures of the impact of training, raises the issue of "cream-skimming" by program operators (Bassi, 1984). Program staff concerned solely with their site's performance relative to the standard should admit into the program applicants who are likely to be employed at good wages (the "cream") regardless of whether or not they benefit from the program. By contrast, they should avoid applicants who are less likely to be employed after leaving training or have low expected wages, even if the impact of the training for such persons is likely to be large. The implications of cream-skimming for equity are clear. If it exists, program operators are directing resources away from those most in need. However, its implications for efficiency depend on the empirical relationship between shortterm outcome levels and longterm impacts. If applicants who are likely to be subsequently employed also are those who benefit the most from the program, performance standards indirectly encourage the efficient provision of training services.[43]

A small literature examines the empirical importance of cream-skimming in JTPA programs. Anderson et al. (1991, 1993) look for evidence of cream-skimming by comparing the observable characteristics of JTPA participants and individuals eligible for JTPA. They report evidence of cream-skimming defined in their study as the case in which individuals with fewer barriers to employment have differentially higher probabilities of participating in training. However, this finding may arise not from cream-skimming by JTPA staff, but because among those in the JTPA eligible population, more employable persons self-select into training.[44]

Two more recent studies address this problem. Using data from the NJS, Heckman and Smith (1998d) decompose the process of participation in JTPA into a series of stages. They find that much of what appears to be cream-skimming in simple comparisons between participants' and eligibles' characteristics is self-selection. For example, high school dropouts are very unlikely to be aware of JTPA and as a result are unlikely ever to apply. To assess the role of cream-skimming, Heckman et al. (1996c) study a sample of applicants from one of the NJS training centers. They find that program staff at this training center do not cream-skim, and appear instead to favor the hard-to-serve when deciding whom to admit into the program. Such evidence suggests that cream-skimming may not be of major empirical importance, perhaps because the social service orientation of JTPA staff moderates the incentives provided by the performance standards system, or

[42] See Heckman and Smith (1998c) and the essays in Heckman (1998b) for more detailed descriptions of the JTPA performance standards system. Similar systems based on the JTPA system now form a part of most US training programs.

[43] Heckman and Smith (1998c) discuss this issue in greater depth. The discussion in the text presumes that the costs of training provided to different groups are roughly equal.

[44] Program staff often have some control over who applies through their decisions about where and how much to publicize the program. However, this control is much less important than their ability to select among program applicants.

because of local political incentives to serve more disadvantaged groups. For programs in Norway, Aakvik (1998) finds strong evidence of negative selection of participants on outcomes. Heinrich (1998) reports just the opposite for a job training program in the United States. At this stage no universal generalization about bureaucratic behavior regarding cream skimming is possible.

Studies based on the NJS also provide evidence on the implications of cream-skimming. Heckman et al. (1997c) find that except for those who are very unlikely to be employed, the impact of training does not vary with the expected levels of employment or earnings in the absence of training. This finding indicates that the impact on efficiency of cream-skimming (or alternatively the efficiency cost of serving the hard-to-serve) is low. Similarly, Heckman et al. (1996c) find little empirical relationship between the outcome measures used in the JTPA performance standards system and experimental estimates of the impact of JTPA training. These findings suggest that cream-skimming has little impact on efficiency, and that administrative performance standards, to the extent that they affect who is served, do little to increase either the efficiency or equity of training provision.

6.6. The conflict between the economic approach to program evaluation and the modern approach to social experiments

We have already noted in Section 5 that under ideal conditions, social experiments identify $\mathrm{E}(Y_1 - Y_0 \mid X, D = 1)$. Without further assumptions and econometric manipulation, they do not answer the other evaluation questions posed in Section 3. As a consequence of the self-selected nature of the samples generated by social experiments, the data produced from them are far from ideal for estimating the structural parameters of behavioral models. This makes it difficult to generalize findings across experiments or to use experiments to identify the policy-invariant structural parameters that are required for econometric policy evaluation.

To see this, recall that social experiments balance bias, but they do not eliminate the dependence between U_0 and D or U_1 and D. Thus from the experiments conducted under ideal conditions, we can recover the conditional densities $f(y_0 \mid X, D = 1)$ and $f(y_1 \mid X, D = 1)$. From non-participants we can recover $f(y_0 \mid X, D = 0)$. It is the density $f(y_0 \mid X, D = 1)$ that is the new information produced from social experiments. The other densities are available from observational data. All of these densities condition on choices. Knowledge of the conditional means

$$\mathrm{E}(Y_0 \mid X, D = 1) = g_0(X) + \mathrm{E}(U_0 \mid X, D = 1)$$

and

$$\mathrm{E}(Y_1 \mid X, D = 1) = g_1(X) + \mathrm{E}(U_1 \mid X, D = 1)$$

does not allow us to separately identify the structure $(g_0(X), g_1(X))$ from the conditional error terms without invoking the usual assumptions made in the non-experimental selec-

tion literature. Moreover, the error processes for U_0 and U_1 conditional on $D = 1$ are fundamentally different than those in the population at large if participation in the program depends, in part, on U_0 and U_1.

For these reasons, evidence from social experiments on programs with different participation and eligibility rules does not cumulate in any interpretable way. The estimated treatment effects reported from the experiments combine structure and error in different ways, and the conditional means of the outcomes bear no simple relationship to $g_0(X)$ or $g_1(X)$ ($X\beta_0$ and $X\beta_1$ in a linear regression setting). Thus it is not possible, without conducting a non-experimental selection study, to relate the conditional means or regression functions obtained from a social experiment to a core set of policy-invariant structural parameters. Ham and LaLonde (1996) present one of the few attempts to recover structural parameters from a randomized experiment, where randomization was administered at the stage where persons applied and were accepted into the program. The complexity of their analysis is revealing about the difficulty of recovering structural parameters from data generated by social experiments.

In bypassing the need to specify economic models, many recent social experiments produce evidence that is not informative about them. They generate choice-based, endogenously stratified samples that are difficult to use in addressing any other economic question apart from the narrow question of determining the impact of treatment on the treated for one program with one set of participation and eligibility rules.

7. Non-experimental evaluations

7.1. The problem of causal inference in non-experimental evaluations

Without invoking the very non-experimental methods they seek to avoid, social experiments cannot address many questions of interest to researchers and policymakers. Even if they could, such data are generally not available. As a result, analysts must rely on "observational" or non-experimental methods to address the problem of selection bias resulting from non-random participation of individuals in employment and training programs.

In an experimental evaluation, information from the control group is used to fill in missing counterfactual data for the treatments. As we have seen, under the assumptions specified in Section 5, an experiment is most successful in generating certain counterfactual means. In a non-experimental evaluation, analysts must replace these missing data with data on non-participants along with assumptions different from those invoked when using the method of social experiments.

To illustrate this point and to highlight an important distinction between experimental and non-experimental solutions to the evaluation problem, consider Fig. 7. It presents a model of potential outcomes in which each outcome takes on one of two possible values. For training participants, Y_1 equals one if the individual is employed after completing

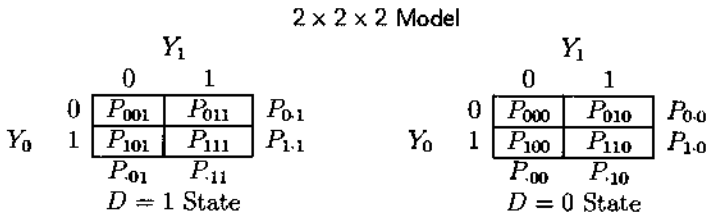

Fig. 7. $2 \times 2 \times 2$ model. Y_1 is an indicator variable for whether or not a person would be employed if trained; Y_0 is an indicator of employment without training. P_{abc} is the probability that $Y_0 = a$, $Y_1 = b$ and $D = c$.

training and equals zero otherwise. For non-participants Y_0 is defined similarly. As before, $D = 1$ for persons who select into training (but who may be excluded in an experimental evaluation) and $D = 0$ otherwise. When program evaluators have access to experimental data, they observe both Y_1 and Y_0 (but never both at the same time for the same person) for persons who select into training. That is, they observe the row and column totals for the $D = 1$ table, but not the proportion of persons for whom $D = 1$ who are in each individual cell. For example, the experimental controls enable the analyst to estimate the proportion of the persons selecting into training ($D = 1$) who would not have been employed in the absence of training, denoted $P_{0\cdot1}$, but not the proportion of persons selecting into training who would not have been employed either with or without training, denoted P_{001}. In order to estimate this proportion, we require another assumption, such as that training did not cause anyone to be non-employed who otherwise would have been employed. This "monotonicity" assumption (training can only make people better off), first invoked in Heckman and Smith (1993), allows us to set $P_{101} = 0$. In that case we can fill in the remaining elements of the table using the row and column totals. The proportion of trainees whose employment status changes as a result of training is now given by P_{011}. When the monotonicity assumption is imposed onto the data from experimental evaluations of training, P_{011} is typically relatively small (see, e.g., Heckman and Smith, 1993). Training causes a relatively small proportion of trainees to switch from the non-employment state to the employment state.

Analysts who have access only to non-experimental data observe only the column totals in the $D = 1$ table and the row totals in the $D = 0$ table. In addition, the proportion of people who take training is known. This can be determined from an experiment that randomizes eligibility but not from an experiment that randomizes among those who apply and are accepted into the program. The remaining elements of both tables, including the other row and column totals, are unknown. The task in observational studies is to find a set of conditioning variables and to impose an appropriate set of assumptions so that the row totals in the $D = 0$ table can be used to estimate the missing row totals in the $D = 1$ table. Regardless of the conditioning variables used or assumptions imposed, there always exists a set of minimal assumptions necessary to identify the impact of training that cannot be tested with the data. The same is true for the analysis of experimental data; the assumptions of no randomization bias or the unimportance of sample attrition cannot be

tested with the data typically generated from experimental evaluations. Both experimental and non-experimental approaches require assumptions that cannot be tested without collecting data specifically designed to test the assumptions of the model.

7.2. *Constructing a comparison group*

All evaluations are based on comparisons between treated and untreated persons. The comparisons may be constructed using the same persons in the treated and untreated states as in the before–after estimator. More commonly, different persons are compared.

The evaluation literature makes an artificial distinction between the task of creating a comparison group and the task of selecting an econometric estimator to apply to that comparison group. In truth, all estimators define an appropriate comparison group and the choice of a comparison group affects the properties of an estimator. The act of constructing or selecting a valid estimator entails assumptions about the samples on which it should be applied.

This simple point is usually overlooked in the empirical literature on program evaluation. It is common to observe analysts first constructing a comparison group on the intuitive principle of making the comparison group "comparable" in some way or other to the treatment group, and then to debate the choice of an estimator as if all estimators defined for random samples of the population can be applied to a comparison group so constructed. Many econometric estimators are only valid for random samples of the population. When non-random samples are generated, the estimators are sometimes no longer valid and have to be modified to account for the impact of the sampling rule used to generate the comparison samples.

The most common instance of this point arises in oversampling participants compared to non-participants. Program records are often abundant for participants; comparison samples often have to be collected at considerable cost. The ratio of program records to comparison group records is usually much larger than one. Simply pooling the two samples misrepresents the population proportion of persons taking training. In order to use the many conventional econometric methods that assume random sampling on such data, the samples have to be reweighted (see the discussion in Heckman and Robb, 1985a, 1986a). A special class of "control function" estimators that we define below does not have to be reweighted. However, instrumental variables estimators have to be reweighted in this case. Different classes of estimators exhibit different degrees of sensitivity to departures from random sampling in constructing comparison groups.

A second example is contamination bias, which we discuss in detail in Section 7.7. Many comparison groups include persons who have actually participated in the program but who have not been recorded as having done so. Again, estimators suitable to random samples without such measurement error on treatment status have to be modified for contaminated samples (Heckman and Robb, 1985a; Imbens and Lancaster, 1996).

A third example concerns the widespread practice of "matching" treatment and comparison group members on dimensions such as pre-program earnings. The literature

often distinguishes between "screening" on characteristics and matching. Screening usually refers to the application of certain broad rules (e.g., income below a certain level) to select observations from a source sample into a comparison sample; matching refers to alignment of trainees and comparison group members over narrower intervals. Both are a form of matching as we define it below and the distinction between them is of no practical value.

More serious are the consequences of this type of matching on the performance of econometric estimators. Matching on variables that are stochastically dependent on the errors of the model sometimes alters the stochastic structure of the errors. Econometric estimators that are valid for random samples can be invalid when applied to the samples generated by matching procedures.

To illustrate the foregoing point, consider the common-coefficient autoregressive estimator introduced into the econometric evaluation literature in Heckman and Wolpin (1976). Using decision rule (6.3) and assuming that agents make their decisions in an environment of perfect certainty and that enrollment into the program only occurs in period k,

$$Y_t = \beta + \alpha D + U_t, \qquad \text{for } t > k, \tag{7.1a}$$

$$Y_t = \beta + U_t, \qquad \text{for } t \leq k, \tag{7.1b}$$

$$U_t = \rho U_{t-1} + \varepsilon_t, \tag{7.1c}$$

where ε_t is an independently and identically distributed error with mean zero. In terms of the model of potential outcomes introduced in Section 3, $Y_t = DY_{1t} + (1 - D)Y_{0t}$ and $Y_{1t} - Y_{0t} = \alpha$, the parameter of interest. The model is in the form of Eq. (3.10) with an autoregressive error. The assumptions about the error terms are typically invoked about random samples of the population. Selection bias in this model arises because of the covariance between D and U_t. In a model with perfect capital markets, only if $\rho=0$ would there be no selection bias.[45]

If we have access to panel data, we can use two post-program observations to estimate α.[46] Write

$$Y_{t-1} = \beta + \alpha D + U_{t-1},$$

where $t - 1 > k$, so that

$$U_{t-1} = Y_{t-1} - \beta - \alpha D.$$

Substituting into (7.1c) and collecting terms, we may rewrite (7.1a) as

[45] However, this result crucially depends on the perfect capital market assumption as we noted in Section 6.3.4

[46] As noted in Heckman and Robb (1985a, 1986a) and below, this estimator can also be applied to repeated cross-section data.

$$Y_t = \beta(1-\rho) + \alpha(1-\rho)D + \rho Y_{t-1} + \varepsilon_t. \tag{7.2}$$

Under decision rule (6.3), D is orthogonal to ε_t even though agents are making their participation decisions under perfect certainty. Least squares applied to (7.2) identifies ρ, and hence α and β. This estimator can be applied to training programs or schooling. Its great advantage is that it can be implemented using only post-program outcome measures provided $\rho \neq 1$. Properties of this estimator are presented in Section 7.6.

Another way to identify α is to use instrumental variables or classic selection bias estimators which we describe in detail below. Assuming random sampling, both of these estimators identify α.

Suppose, however, that we first "match" on pre-training earnings, $Y_{t'}, t' < k$, in order to construct a comparison sample of non-participants. Consider a simple screening rule: select observations into the sample if $Y_{t'} < l$. This rule is widely used in constructing comparison samples. How are the error structure (7.1c) and the properties of the three estimators just discussed affected by the application of such screening rules? The autoregressive estimator just presented using post-program observations is unaffected by these sampling rules. It continues to identify α and ρ. This is immediately seen because $\mathrm{E}(\varepsilon_t \mid D = 1, Y_{t-1}, Y_{t'} < l) = 0$ since ε_t is independent of $Y_{t'}, t' < k$, and Y_k.

However, matching affects the distribution of the errors. This makes a sample selection model based on a distributional assumption appropriate to a random sample inappropriate when applied to a matched sample. In this case, two selection rules generate the outcomes; classical selection estimators that only account for agent self-selection do not account for the selection bias induced by the analysts' matching procedure. Instrumental variables methods appropriate to random samples in general become inconsistent when applied to matched samples for reasons exposited in Section 7.7.

Another strategy for defining a comparison group is to use program applicants who drop out of the application and enrollment process before receiving training. Such comparison groups include persons who applied and were rejected from the program, those who were admitted but never showed up for training ("no-shows"), or early program dropouts. (No-shows are used in, e.g., Cooley et al., 1979; LaLonde, 1984; Bell et al., 1995; Heckman et al., 1997a). In samples based on no-shows, two decision rules – whether or not to apply to the program and whether or not to stay in the program if accepted – determine which non-participants end up in the comparison group sample. The properties of econometric estimators have to be examined to see if they are robust to such sample selection rules. Analytically, this is the same problem as arises in the construction of matched samples, except that in this case the decision rules of agents govern the construction of samples. Estimators valid for samples generated by one decision rule need not be valid for another.

A brief summary of the screening and matching criteria used in several major evaluations is presented in the last row of Tables 5 and 6. Table 7, based on Barnow (1987), presents a more exhaustive list of characteristics used to match and control for differences in evaluations of the US CETA program, the immediate predecessor of JTPA. Combining

matching and different non-experimental evaluation methods that break down when applied to matched samples constitutes an important source of variability across these studies, one that has more to do with the properties of the estimators selected than with the properties of the programs being studied.

In the literature, the act of specifying a comparison group and then making conditional mean comparisons between participants and comparisons is equivalent to defining a matching estimator. The matching estimator may be embellished by further adjustments as we note below. A different comparison group might be specified for each treatment observation. The potential sample from which the comparison group is taken includes all persons who do not take treatment. Further restrictions on this universe define different matching rules.

7.3. Econometric evaluation estimators

All evaluation estimators are based on the three basic estimation principles introduced in Section 4. They entail making some comparison of treated individuals with the untreated. The comparison may be between treated and untreated persons at a point in time as in the cross-section estimator; it may be between the same persons in the treated and untreated states as in the before–after estimator; or it may be a hybrid of the two principles as in the difference-in-differences estimator. In this section, we extend these basic estimators to allow for conditioning variables and to exploit knowledge of the serial correlation properties of error terms.

The estimators within each class differ in the way they adjust, condition or transform the data in order to construct the counterfactual $E(Y_0 \mid X, D = 1)$. Throughout the rest of this section, we consider how the various estimators construct the counterfactual and what assumptions they make about individual decision processes that determine program participation. We motivate this discussion using the simple decision and outcome models of Section 6.3. The first class of estimators that we consider are cross-section estimators based on matching methods. These estimators are frequently used in studies by consulting firms because they are relatively easy to explain to their clients. A disadvantage of this approach is that it requires strong underlying assumptions about the selection process into training. Although the method is usually applied in a cross-sectional setting, matching can be generalized to apply to panel settings as in Heckman et al. (1997a, 1998c). The second class of cross-section methods we consider are selection bias correction methods developed in Heckman (1976, 1979) or Heckman and Robb (1985a, 1986a). This approach is often used in studies of European training programs. It too can be extended to apply to panel data, but is most frequently applied in a cross-sectional setting.

Program evaluations by academic labor economists in the United States have relied almost exclusively on a third class of estimators: longitudinal methods that extend the before–after and difference-in-differences estimators. An implicit belief shared by the authors of these studies is that longitudinal methods are more robust than cross-section selection bias correction methods, which are sometimes dismissed as being "functional

Table 5
Explanatory variables used in previous studies[a]

	MDTA data		CLMS-based studies[b]	
	Ashenfelter (1978)	Ashenfelter and Card (1985)	Dickinson et al. (1987)	Bryant and Rupp (1987)
Program, year, outcome variable	MDTA classroom trainees enrolled in first 3 months of 1964; 1965–1969 annual social security earnings	1976 CETA trainees, 1977 and 1978 annual social security earnings	CETA trainees enrolled in 1976; 1978 annual social security earnings	Two cohorts of CETA trainees; 1977 and 1978 annual social security earnings
Local labor market information	No	No	No	No
Age, race, sex	Yes	Yes	Yes	Yes
Education	No	Yes	Yes	Yes
Training history	No	No	No	No
Children	No	No	No	Yes
Employment histories	No	No	Yes (recent)	Yes (recent)
Hours worked	No	Yes	Yes	Yes
Unemployment histories	No	No	Yes (recent)	Yes (recent)
Welfare receipt	No	No	Yes	No

Earnings histories	5 years pre-program 5 years post-program	2 years pre-program 2 years post-program	Same as Ashenfelter and Card (1985)	4 years pre-enrollment
Matching criteria[c]	None specified	(a) 1975 earnings $\leq$ \$20K and household income $\leq$ \$30K (b) In labor force in March, 1976 (c) Age greater than 20	Matching based on a metric over a vector of predictors of 1978 earnings including lagged earnings (1970–1975), unemployment in 1975, worked in public sector, in labor force in March 1976 and demographics.	Matching on 1976 earnings, change in earnings 1975–1976, 1975 labor force status, family income, in labor force in 1975 or at March 1976 interview and demographics.

[a] Source: Heckman et al. (1997a, Table 1). Notes: "Used" means employed in the matching process and/or included in the outcome equation.

[b] CLMS, CETA Longitudinal Manpower Survey. The CLMS data matched Social Security longitudinal earnings records to CETA participants and CPS comparison group members from the March 1976 and 1977 CPS. All of the CLMS-based studies use the social security earnings data except for Bassi (1983), which also uses the CPS earnings data. All of the personal and family information available in the CPS, including shortterm employment and labor force participation histories, are available in the CLMS but not necessarily used in the analyses based upon it.

[c] "Matching Criteria" indicate the criteria for membership in the comparison group. This is sometimes referred to as "screening" in the literature.

Table 6
Explanatory variables used in previous studies[a]

	CLMS-based studies[b]	NSW (Supported Work) data		NJS data
	Bassi (1983, 1984)	LaLonde (1986)	Fraker and Maynard (1987)	Heckman et al. (1998b)
Program and outcome variable	CETA, 1977 and 1978 annual social security earnings	NSW, 1979 annual social security and PSID survey earnings	NSW, 1977, 1978 and 1979 annual earnings for AFDC women and youth	NJS, monthly survey earnings in the 18 months after random assignment or measured eligibility
Local labor market information	No	No	No	Yes
Age, race, sex	Yes	Yes	Yes	Yes
Education	?[d]	Yes	Yes	Yes
Training history	No	No	No	Yes (partial)
Children	?	Yes	Yes	Yes
Employment histories	?	No	No	Yes
Hours worked	?	No	No	Yes
Unemployment histories	?	No	No	Yes (1 year)
Welfare receipt	?	Yes	Yes	Yes (1 year)
Earnings histories	Same as Bryant and Rupp (1987)	Two years pre-program Two years post-program	Two years pre-program Two years post-program	Five years pre-program Two years post-program

Matching criteria[c] (criteria for membership in comparison sample)	Same as Bryant and Rupp (1987); also uses a random sample from the March 1976 CPS	PSID: household head in 1975–1979; CPS: March 1976 CPS earnings matched to SSA earnings; screens based on 1976 personal and household income	Three samples: I: eligible in sample period; screen out in-school youth; AFDC women match on age of youngest child and welfare receipt II: eligible in sample period; cell matching based on predictors of 1979 SSA earnings including prior earnings, change in earnings, education, family income and demographics III: eligible in sample period; match on earnings estimated on eligible non-participant sample, age and sex	Within age and sex groups, match on propensity score based on site, race, age, schooling, marital status labor force status history, number of recent jobs, training history, household size and recent earnings.

[a] Source: Heckman et al. (1997a, Table 1). Notes: "Used" means employed in the matching process and/or included in the outcome equation.

[b] CLMS, CETA Longitudinal Manpower Survey. The CLMS data matched Social Security longitudinal earnings records to CETA participants and CPS comparison group members from the March 1976 and 1977 CPS. All of the CLMS-based studies use the social security earnings data except for Bassi (1983), which also uses the CPS earnings data. All of the personal and family information available in the CPS, including shortterm employment and labor force participation histories, are available in the CLMS but not necessarily used in the analyses based upon it.

[c] "Matching Criteria" indicate the criteria for membership in the comparison group. This is sometimes referred to as "screening" in the literature.

[d] "?" indicates that the study does not specify the variables used.

Table 7
Matching characteristics used in CETA evaluation studies[a]

	Westat (1981)	Westat (1984)	Bassi (1983)	Bassi et al. (1984)	Bloom and McLaughlin (1982)	Dickinson et al. (1984)	Geraci (1984)
Program entry	7/75–6/76	7/75–6/76 (A) 7/75–6/77 (B)	7/75–6/76	7/76–9/77	1/75–6/76	1/76–12/76	7/75–7/76
Post-program period	1977	1977 (A) 1978 (B)	1977, 1978	1978, 1979	1976, 1977, 1978	1978	1977–1979 average
CETA participants included in analysis	Ages 14–16 and enrolled in CT, PSE, OJT, WE, or multiple training types; over 7 days in program; prior year earnings less than 20,000; prior year family income less than 30,000; terminated from program by 12/76; valid SSA match on 3 of 5 criteria	Same as Westat (1981) except family income excludes participant's earnings	Same as Westat (1981)	Welfare recipients and other economically disadvantaged persons ages 18–65 and youth ages 13–22; no other restrictions	Ages 25–60 and enrolled in CT, OJT, or WE only. Must have over 7 days in program	Ages 16–64 and not in summer youth program; must be complete or close SSA match and not in program in 1978	Same as Westat (1981) except only persons over age 22
CPS individuals eligible for comparison group	Same age, earnings, income and SSA match: in labor force in 3/76 or worked in 1975	Same as Westat (1981).	Same as Westat (1981).	For ages 18–65, must be on welfare or economically disadvantaged: for youths 13–22, same as Westat (1984b) (B)	Ages 25–60; earned less than SSA maximum from 70–75; 1975 family income < $30,000	Adults in labor force in 3/76; youth in labor force in 3/76 or worked in 1975	Same as Westat (1984) (A) group

Matching procedure	Cell matching for 1972–1974 earnings groups; for low earners: exact match on sex, race, and age; collapsing permitted on education, family income, labor force experience, family head status, 1975 SSA earnings, change in SSA earnings 74–75, change in SSA earnings 73–74, poverty status and private sector employment; for intermediate earners: exact match on sex, race, 1975 SSA earnings, change in SSA earnings; collapsing permitted on other variables; for high earners: same as intermediate earners except family income	Cell matching for each activity; exact match on sex, 1975 SSA earnings for (A) or 1976 SSA earnings for (B), change in SSA earnings for two previous years (73–74 and 74–75 or 74–75 and 75–76) and race; collapsing permitted on age, education, family income, prior year labor force experience, family head status, and poverty status	Same as Westat (1981)	All economically disadvantaged persons and welfare recipients ages 18–65 included in adult study. For youth ages 13–22, used Westat (1981) youth matched comparison groups	All CPS individuals who met the above criteria were included	Weighted nearest-neighbor match based on SSA earnings in 1970–1975, square of 1975 SSA earnings, race and ethnicity, age, age^2, age^3, family head status, occupational categories, public sector employment, poverty status, AFDC receipt, UI receipt, percent of time worked in 1975, percent of time worked in 1974, CPS reported earnings for those at SSA maximum, and 15 interaction variables; match groups formed overall and by activity	Same as Westat (1981) (A)

Table 7 (*continued*)

	Westat (1981)	Westat (1984)	Bassi (1983)	Bassi et al. (1984)	Bloom and McLaughlin (1982)	Dickinson et al. (1984)	Geraci (1984)
	given less priority in cell collapsing						
Regression procedure	Weighted least squares with separate regressions for each race-sex-earnings group	Weighted least squares with separate regressions for each activity	Difference-in-differences with separate regressions by sex-race group	Difference-in-differences with separate regressions by sex-race-welfare group	OLS with fixed effects, individual time trends and correction for earnings drop for participants	Ordinary least squares with separate regressions by age-sex-activity group	Two-step procedure: (1) probit for positive earnings; (2) weighted least squares for positive earners with separate analysis by sex
Regressors	Family head status, education, prior work in private sector, 1973 SSA earnings, 1974 SSA earnings, proxy for cyclical unemployment, family income, prior labor force status, age, educational disadvantage and status (ages 16–18 only) veteran	Same as Westat (1981)	Age, age^2	Age, age^2	Age, age^2, education, education^2, family size, minority status, head of household status, current marital status, past marital status, presence of children under age 4, presence of children ages 7–18	Same regressors as used for matching	Age, age^2, education, marital status, head of household status, economically disadvantaged status, minority status, presence of children under age 6 (females only), presence of children ages 6–18 (females only), interaction terms for experience

status (males only), presence of children under age 6 (females only), presence of children ages 6–18 (females only)		and education

[a] Source: Barnow (1987, Table 2).

form dependent." However, we demonstrate below that as currently utilized in the applied evaluation literature, longitudinal estimators depend on functional form assumptions. Moreover, longitudinal estimators are often much less robust to choice-based sampling and other matching and screening procedures used to produce comparison samples in the empirical literature than are cross-section sample selection estimators. In the remainder of this section, we discuss the identifying assumptions that underlie the main methods used in evaluation research, and sketch out how they are implemented to produce practical estimators.

We remind the reader that throughout this chapter, we use X variables that are not determined by D. Letting X be the vector of conditioning variables and Y^P a vector of potential outcomes, we write $Y_t^P = (Y_{0t}, Y_{1t})$, and $Y^P = (Y_1^P, \ldots, Y_T^P)$, $X = (X_1, \ldots, X_T)$. We define the admissible X on which we condition to define parameters as those X that satisfy

$$f(X \mid D, Y^P) = f(X \mid Y^P). \tag{7.A.1}$$

where $f(X \mid D, Y^P)$ is the density of X given D and Y^P and $f(X \mid Y^P)$ is the density of X given Y^P.[47] This assumption says that given the potential outcomes in both states, the actual occurrence of D provides no more information on X ("Does not cause X"). We maintain this assumption in order to avoid masking the effects of D on outcomes by conditioning on variables that are determined by D. Other definitions are possible but we maintain this one to make our analysis interpretable and to avoid certain technical problems in making forecasts with our parameters. Heckman (1998a) presents a more extensive discussion of this condition and relates it to definitions of causality and exogeneity in the econometric time series literature.

7.4. Identification assumptions for cross-section estimators

When participation in training is voluntary, and evaluators have access to cross-sectional data, they can construct the distribution of outcomes for participants, $F(Y_1 \mid X, D = 1)$, and for non-participants, $F(Y_0 \mid X, D = 0)$. They use $F(Y_0 \mid X, D = 0)$ to approximate $F(Y_0 \mid X, D = 1)$, which runs the risk of selection bias. When using this approximation, the bias in estimating $\mathrm{E}(Y_1 - Y_0 \mid X, D = 1)$ is given by

$$B(X) = \mathrm{E}(Y_0 \mid X, D = 1) - \mathrm{E}(Y_0 \mid X, D = 0). \tag{7.3}$$

Many schemes have been proposed to circumvent this bias. We begin by considering the intuitively-appealing method of matching.

7.4.1. The method of matching

The method of matching assumes that analysts have access to a set of conditioning variables, X, such that, within each "stratum" defined by X, the counterfactual outcome distribution of the participants is the same as the observed outcome distribution of the non-

[47] Heckman and Borjas (1980) develop this non-causality condition.

participants.[48] The statistical matching literature assumes access to a set of X variables such that

$$(Y_0, Y_1) \perp\!\!\!\perp D \mid X, \tag{7.4}$$

where "$\perp\!\!\!\perp$" denotes independence and X denotes variables on which conditioning is conducted. As a consequence of (7.4), the distributions of outcomes $F(Y_0 \mid D = 1, X) = F(Y_0 \mid D = 0, X) = F(Y_0 \mid X)$ and $F(Y_1 \mid D = 1, X) = F(Y_1 \mid D = 0, X) = F(Y_1 \mid X)$. The method appeals to the intuitive principle that it is possible to "adjust away" differences between participants and non-participants using the available regressors.

If assumption (7.4) is valid we can use non-participants to measure what participants would have earned had they not participated, provided we condition on the variables X. To ensure that this assumption has empirical content, it also is necessary to assume that there are participants and non-participants for each X for which we seek to make a comparison. More formally, this means that

$$0 < \Pr(D = 1 \mid X) < 1 \tag{7.5}$$

over the set of X values where we seek to make a comparison. To satisfy this condition, at least in large samples, there must be both participants and non-participants for each X. In a finite sample of any size, we replace this condition by the empirical probability.[49] This condition ensures that the distributions in (7.4) are defined for all X that satisfy it. As we demonstrate below in Section 8.2, this assumption has important practical consequences for training evaluations. Failure to satisfy this condition appears to be one major reason why matching methods produce biased estimates of the impact of training in the NJS study. The treatment parameter $\mathrm{E}(Y_1 - Y_0 \mid X, D = 1)$ cannot be identified for values of X where (7.5) is violated.

Under assumptions (7.4) and (7.5), matching produces a comparison group that resembles an experimental control group in one key respect: conditional on X, the distribution of the counterfactual outcome, Y_0, for the participants is the same as the observed distribution of Y_0 for the comparison group. In particular, as long as the means exist, assumptions (7.4) and (7.5) imply that

$$\mathrm{E}(Y_0 \mid X, D = 1) = \mathrm{E}(Y_0 \mid X, D = 0), \tag{7.6a}$$

and that

$$\mathrm{E}(Y_1 \mid X, D = 1) = \mathrm{E}(Y_1 \mid X, D = 0). \tag{7.6b}$$

Therefore, for each point in X, the bias $B(X) = 0$. However, this assumption does not imply no selection bias, i.e., that $\mathrm{E}(U_0 \mid X, D = 1) = 0$. Instead, like experiments, match-

[48] The first published instance of the use of this method of which we are aware is Fechner (1860).

[49] The support of X consists of those values of X with positive density. Assumptions (7.4) and (7.5) are called "strong ignorability" by Rosenbaum and Rubin (1983).

ing balances the bias:

$$\mathrm{E}(U_0 \mid X, D = 1) = \mathrm{E}(U_0 \mid X, D = 0) = \mathrm{E}(U_0 \mid X). \tag{7.7}$$

In an ideal experiment, we obtain a comparison group via randomization among persons for whom $D = 1$. Matching emulates an experiment by replacing randomization with conditioning on a set of X variables. Conditional on those values, persons randomly select into the program. There are no selective differences in Y_0 outcomes between participants and non-participants given X. Randomization at the stage where persons enter the program also may be thought of as a form of conditioning (Heckman, 1996). It operates conditional on $D = 1$. Under the conditions that justify it, randomization generates a control group for each X in the participant population. Similarly, under assumption (7.4), matching generates a comparison group, but only for these X values that satisfy (7.5), which in practice is often a much smaller set of values than would be the case with randomization.

In Section 8.2 below, we draw on the work of Heckman et al. (1998b) and demonstrate that the reduction in the set of X for which the parameter of interest is defined can be substantial. Further, because the impact parameter may depend on X, the parameter estimated by an experimental evaluation and the parameter estimated by matching may be different.

When the Rosenbaum–Rubin assumptions (7.4) and (7.5) are invoked, it is possible to construct both the "treatment on the treated" parameter $\mathrm{E}(Y_1 - Y_0 \mid X, D = 1)$ and the effect of "non-treatment on the non-treated" $\mathrm{E}(Y_0 - Y_1 \mid X, D = 0)$. Only assumption (7.6a) is required if we are interested in the mean effect of treatment on the treated. It permits agents to select into the program on the basis of U_1 but not U_0. Assuming that $\mathrm{E}(U_0 \mid X) = 0$, it implicitly defines the parameter "treatment on the treated" in an asymmetric way:

$$\mathrm{E}(Y_1 - Y_0 \mid X, D = 1) = \mu_1(X) - \mu_0(X) + \mathrm{E}(U_1 \mid X, D = 1)$$

because $\mathrm{E}(U_0 \mid X, D = 1) = \mathrm{E}(U_0 \mid X) = 0$. This parameter no longer equals the effect of treatment on a randomly selected person as it would if (7.4) held. Assumption (7.6b) allows us to identify the mean effect of non-treatment on the non-treated.

Using representation (3.1a) and (3.1b), (7.4) and (7.5) imply that $\mathrm{E}(U_0 \mid X, D = 1) = \mathrm{E}(U_0 \mid X, D = 0) = \mathrm{E}(U_0 \mid X) = 0$ and $\mathrm{E}(U_1 \mid X, D = 1) = \mathrm{E}(U_1 \mid X, D = 0) = \mathrm{E}(U_1 \mid X) = 0$. Thus conditioning on X, the two parameters "treatment on the treated" and "the effect of randomly assigning a person with characteristics X to the program" are the same.[50] From an economic standpoint, assumption (7.4) rules out selection into the program on the basis of unobservables (U_0, U_1) that may be partially known to people taking training but are unknown to the observing economist. In terms of the random coefficient model of Section 3, it rules out correlation between D and the difference in unobserved components, $(U_1 - U_0)$. It defines an implicit economic model that assumes that agents do not enter the program on the basis of gains unobserved by analysts. Thus it is

[50] This is also true if $Y_1 = g_1(X) + U_1$ and $Y_0 = g_0(X) + U_0$ and $\mathrm{E}(U_1 \mid X) \neq 0$ and $\mathrm{E}(U_0 \mid X) \neq 0$. In that case, $\mathrm{E}(Y_1 - Y_0 \mid X, D = 1) = g_1(X) - g_0(X) + \mathrm{E}(U_1 - U_0 \mid X)$ so that the two parameters are the same.

a method congenial with the assumption that α in (6.3) is a common coefficient, or that if α varies among persons with identical X, then participation in the program is not based on this variation. In the context of that model, the "cost of participation" or any of the variables generating participation, but not outcomes, are valid conditioning variables. Thus, if the costs of participation are distributed independently of all other variables and if Y_{0k} is independent of Y_{0l}, then conditioning on c or Y_{0k} will satisfy the conditions required to justify the matching estimator. However, as we explained in Section 6.3.1, if we condition on both cost of participation and Y_{0k}, we violate condition (7.5). Matching breaks down if there is too much information and other methods must be used to evaluate the program.[51]

To operationalize the method of matching, assume two samples: "t" for treatment and "c" for comparison group. Unless otherwise noted, observations are statistically independent. Simple matching methods are based on the following idea: For each person i in the treatment group, we find some group of "comparable" persons. The same individual may be in both groups if that person is treated at one time and untreated at another. We denote outcomes in the treatment group by Y_i^t and we match these to the outcomes of a subsample of persons in the comparison group to estimate a treatment effect. In principle, we can use a different subsample as a comparison group for each person.

In practice, we can construct matches on the basis of a neighborhood $C(X_i)$, where X_i is a vector of characteristics for person i. Neighbors to treated person i are persons in the comparison sample whose characteristics are in neighborhood $C(X_i)$. Suppose that there are N_c persons in the comparison sample and N_t in the treatment sample. Thus the persons in the comparison sample who are neighbors to i, are persons j for whom $X_j \in C(X_i)$, i.e., the set of persons $A_i = \{j \mid X_j \in C(X_i)\}$. Let $W(i,j)$ be the weight placed on observation j in forming a comparison with observation i and further assume that the weights sum to one,

$$\sum_{j=1}^{N_c} W(i,j) = 1,$$

and that $0 \leq W(i,j) \leq 1$. Then we form a weighted comparison group mean for person i, given by

$$\bar{Y}_i^c = \sum_{j=1}^{N_c} W(i,j) Y_j^c, \tag{7.8}$$

and the estimated treatment effect for person i is $Y_i - \bar{Y}_i^c$.

Heckman et al. (1997a) survey a variety of alternative matching schemes proposed in the literature. Here we briefly introduce two widely used methods. The nearest-neighbor

[51] The regression discontinuity design estimator discussed in Section 7.4.6 can be applied here as a limit form of the matching estimator that identifies $\mathrm{E}(Y_1 - Y_0 \mid X, D = 1)$ at one point.

matching estimator defines A_i such that only one j is selected so that it is closest to X_i in some metric:

$$A_i = \{j \mid \min_{j \in \{1,\ldots,N_c\}} \| X_i - X_j \| \},$$

where $\| \ \|$ is a metric measuring distance in the X characteristics space. The Mahalanobis metric is one widely used metric for implementing the nearest-neighbor matching estimator. The metric used to define neighborhoods for i is

$$\| \ \| = (X_i - X_j)' \Sigma_c^{-1} (X_i - X_j),$$

where Σ_c is the covariance matrix in the comparison sample. The weighting scheme for the nearest neighbor matching estimator is

$$W(i,j) = \begin{cases} 1 & \text{if } j \in A_i, \\ 0 & \text{otherwise.} \end{cases}$$

A version of nearest-neighbor matching, called "caliper" matching (Cochran and Rubin, 1973), makes matches to person i only if

$$\| X_i - X_j \| < \varepsilon,$$

where ε is a pre-specified tolerance. Otherwise person i is bypassed and no match is made to him or her.

Kernel matching uses the entire comparison sample, so that $A_i = \{1, \ldots, N_c\}$, and sets

$$W(i,j) = \frac{K(X_j - X_i)}{\sum_{j=1}^{N_c} K(X_j - X_i)}, \tag{7.9}$$

where K is a kernel. In practice, kernels are typically a standard distribution function such as that for the normal. Kernel matching is a smooth method that reuses and weights the comparison group sample observations differently for each person i in the treatment group with a different X_i. Kernel matching can be defined pointwise at each sample point X_i or for broader intervals.

The impact of treatment on the treated is estimated by forming the mean difference across the i

$$m = \frac{1}{N_t} \sum_{i=1}^{N_t} (Y_i^t - \bar{Y}_i^c) = \frac{1}{N_t} \sum_{i=1}^{N_t} (Y_i^t - \sum_{j=1}^{N_c} W(i,j) Y_j^c). \tag{7.10}$$

We can define this mean for various subsets of the treatment sample defined in various ways. More efficient estimators weight the observations accounting for the variance (Heckman et al., 1997a, 1998c; Heckman, 1998a).

Regression-adjusted matching, proposed by Rubin (1979) and clarified in Heckman et

al. (1997a, 1998c), uses regression-adjusted Y_i, denoted by $A(Y_i) = Y_i - X_i\beta$, in place of Y_i in the preceding calculations. (See the cited papers for the econometric details of the procedure). Regression-adjusted matching methods were widely used in the controversial CETA evaluations conducted in the early 1980s, which we discuss below.

The essence of the idea justifying matching is that conditioning on X eliminates selection bias. Like social experiments, the method requires no functional form assumptions for outcome equations. If, however, a functional form assumption is maintained, as in the econometric procedure proposed by Barnow et al. (1980), it is possible to implement the matching assumption using regression analysis. Suppose that Y_0 is linearly related to observables X and an unobservable U_0, so that $\mathrm{E}(Y_0 \mid X,\ D = 0) = X\beta + \mathrm{E}(U_0 \mid X,\ D = 0)$, and $\mathrm{E}(U_0 \mid X,\ D = 0) = \mathrm{E}(U_0 \mid X)$ is linear in X. Under these assumptions, controlling for X via linear regression allows one to identify $\mathrm{E}(Y_0 \mid X,\ D = 1)$ from the data on non-participants. Such functional form assumptions are not strictly required to implement the method of matching. Moreover, in practice, users of the method of Barnow et al. (1980) do not impose the common support condition (7.5) for the distribution of X when generating estimates of the training effect. The distribution of X may be very different in the trainee ($D = 1$) and comparison group ($D = 0$) samples, so that comparability is only achieved by imposing linearity in the parameters and extrapolating over different regions.

One advantage of the method of Barnow et al. (1980) is that it uses data parsimoniously. If the X are high dimensional, the number of observations in each cell when matching can get very small. Another solution to this problem that reduces the dimension of the matching problem without imposing arbitrary linearity assumptions is based on the probability of participation or the "propensity score," $P(X) = \Pr(D = 1 \mid X)$. Rosenbaum and Rubin (1983) demonstrate that if assumptions (7.4) and (7.5) hold, then

$$(Y_1, Y_0) \perp\!\!\!\perp D \mid P(X) \text{ for } X \in \chi_c, \tag{7.11}$$

for some set χ_c where it is assumed that (7.5) holds in the set. Conditioning on $P(X)$ rather than on X produces conditional independence. Condition (7.11) has the important implication that to construct the desired counterfactual conditional mean $\mathrm{E}(Y_0 \mid P(X), D = 1)$, we require only that

$$B(P(X)) = \mathrm{E}(Y_0 \mid P(X), D = 1) - \mathrm{E}(Y_0 \mid P(X), D = 0) = 0. \tag{7.12}$$

We also could invoke (7.12) in place of (7.11) to define the conditions required to justify matching to estimate mean impacts. Conditioning on $P(X)$ sets $B(P(X)) = 0$ and reduces the dimension of the matching problem down to matching on the scalar $P(X)$. The analysis of Rosenbaum and Rubin (1983) assumes that $P(X)$ is known rather than estimated. Heckman et al. (1998c) present the asymptotic distribution theory for the kernel matching estimator in the cases in which $P(X)$ is known and in which it is estimated both parametrically and non-parametrically. They also answer the question, "If $P(X)$ were known would we match on it or on X?" Using the variance of the estimated average impacts as the choice criterion, the answer is "it depends".

A major advantage of the method of randomized trials over the method of matching is

that randomization works for any choice of X. In the method of matching, there is the same uncertainty about which X to use as there is in the specification of conventional econometric models. Even if one set of X values satisfies condition (7.11) or (7.12), an augmented or reduced version of this set may not. Heckman et al. (1997a) discuss tests that can help determine the appropriate choice of X variables. Any convincing application of the method of matching requires a demonstration that an adequate model for $P(X)$ has been selected. Heckman et al. (1998b) discuss this problem in depth. In the statistics literature, there is no discussion of the choice of X. Implicitly, the advice given there is to use all available regressors. One general rule, already noted in the introduction to this section, is to include in X only variables that are not caused by D given the unobservables. Intuitively, conditioning on variables caused by D masks the true effect of D on outcomes.

The method of matching is sometimes used to estimate $\mathrm{E}(Y_1 - Y_0 \mid X, D = 1)$ at points of $X = x$. More commonly, an averaged version of this parameter is estimated over a set $S(X)$:

$$\mathrm{E}(Y_1 - Y_0 \mid D = 1) = \frac{\int_{S(X)} \mathrm{E}(Y_1 - Y_0 \mid X, D = 1)\mathrm{d}F(X \mid D = 1)}{\int_{S(X)} \mathrm{d}F(X \mid D = 1)}. \tag{7.13}$$

The distinction between the average parameter and the pointwise parameter is an important one. Even though the behavioral motivation and the identifying assumptions are different, it turns out that both the matching estimator and the classical selection estimator can identify (7.13) under very different behavioral assumptions. We now turn to consider the classical selection estimator.

7.4.2. Index sufficient methods and the classical econometric selection model

The most troubling feature of the method of matching is the assumption that selection into a program does not occur on the basis of unobservable (to the economist) gains from the program (U_0 if (7.6a) is assumed; $U_1 - U_0$ if (7.4) is assumed). Depending on the quality of the data at the analyst's disposal, it may or may not be attractive to assume that the analyst knows as much as the people being studied. The method of matching is not robust to violations of this assumption.

The traditional econometric approach to the selection problem adopts a more conservative approach and allows for selection on unobservables. As currently formulated, it assumes an additively separable model relating outcomes to regressors and additive errors, but does not require the strong behavioral assumptions that justify matching. Thus it trades a behavioral assumption for an additive separability assumption. It allows for selection into the program on the basis of unobserved components of outcomes. This approach is in the spirit of much econometric work that builds models to estimate a variety of counterfactual states, rather than just the single counterfactual state required to estimate the mean impact of treatment on the treated, which is the parameter of interest in most evaluations based on the methods of matching or random assignment.

In the simplest econometric approach, two functions are postulated: $Y_1 = g_1(X, U_1)$ and

$Y_0 = g_0(X, U_0)$, where U_0 and U_1 are unobservables. A selection equation is specified to determine which outcome is observed. Separability between X and (U_0,U_1) is assumed, so that

$$Y_1 = g_1(X) + U_1 \qquad \text{and} \qquad Y_0 = g_0(X) + U_0,$$

where for simplicity we assume that $\mathrm{E}(U_1 \mid X) = \mathrm{E}(U_0 \mid X) = 0$ so that $g_1(X) = \mu_1(X)$ and $g_0(X) = \mu_0(X)$. These exogeneity assumptions are not strictly required but for simplicity we maintain them.[52] This assumption *defines* functions called "structural equations" that do not depend on unobserved variables. In this notation, the treatment on the treated parameter is

$$\mathrm{E}(Y_1 - Y_0 \mid X, D = 1) = g_1(X) - g_0(X) + \mathrm{E}(U_1 - U_0 \mid X, D = 1),$$

which combines "structure" and "error" in a somewhat unusual way.

Much applied econometric work is devoted to eliminating the mean effect of unobservables on estimates of functions like g_0 and g_1. However, as previously noted, the mean difference in unobservables is an essential component of the definition of the parameter of interest in evaluating social programs. In the conventional framework, the selection bias that arises from using a non-experimental comparison group is given by

$$B(X) = \mathrm{E}(U_0 \mid X, D = 1) - \mathrm{E}(U_0 \mid X, D = 0).$$

In the standard evaluation problem, the goal is to set $B(X) = 0$, *not* to eliminate dependence between (U_0,U_1) and X and D.

The conventional econometric approach for addressing selection bias partitions the observed variables X into two not necessarily disjoint sets (Q,Z) corresponding to those variables in the outcome equations and those variables in the participation equation, and then postulates exclusion restrictions. It assumes that certain variables appear in Z but not in Q. The conventional approach further restricts the model so that the bias $B(X)$ only depends on Z through a scalar index. Recall that such exclusion restrictions are not required to justify matching as an estimator.[53]

The latent index model of program participation introduced in Section 6 motivates the characterization of selection bias as a function of a scalar index. In that model, we defined the index $IN = H(Z) - V$, where $H(Z)$ is the mean difference in utilities or discounted earnings between the training and non-training states, and V is assumed to be independent of Z. The training indicator, D, then equals one when $IN > 0$ and equals zero otherwise, resulting in $\Pr(D = 1 \mid Z) = F_V(H(Z))$. The conventional econometric selection model further assumes that the dependence of D and the unobservables U_0 and U_1 arises only

[52] Thus we could instead postulate instruments Z such that $\mathrm{E}(U_1 \mid X) \neq 0$ and $\mathrm{E}(U_0 \mid X) \neq 0$ but $\mathrm{E}(U_1 \mid X, Z) = 0$ and $\mathrm{E}(U_0 \mid X, Z) = 0$ in order to define the g_0 and g_1 functions.

[53] Heckman et al. (1997a, 1998c) extend the theory of matching to consider separable models with exclusion restrictions and discuss the efficiency gains from using such restrictions. Exclusion restrictions are natural in the context of panel data models where the variables in the outcome equation are measured in periods after the decision to participate in the program is made.

through V and that Q and Z are independent of U_0 and U_1. These assumptions imply the following:

$$\mathrm{E}(U_0 \mid Z, Q, D = 1) = \mathrm{E}(U_0 \mid V < H(Z)),$$

$$\mathrm{E}(U_0 \mid Z, Q, D = 0) = \mathrm{E}(U_0 \mid V \geq H(Z)),$$

$$\mathrm{E}(U_1 \mid Z, Q, D = 1) = \mathrm{E}(U_1 \mid V < H(Z)),$$

and

$$\mathrm{E}(U_1 \mid Z, Q, D = 0) = \mathrm{E}(U_1 \mid V \geq H(Z)),$$

We could just as well postulate this representation as the starting point for our analysis of the selection estimator. Both the bias, $B(Z)$ and the mean gain of the unobservables, $\mathrm{E}(U_1 - U_0 \mid Z, Q, D = 1)$, depend on Z only through the index $H(Z)$. When F_V is assumed to be strictly monotonic almost everywhere, we may write $H(Z) = F_V^{-1}(\Pr(D = 1 \mid Z))$ and the bias and mean gain terms depend on Z solely through the probability of participation P. The bias is now given by

$$B(P(Z)) = \mathrm{E}(U_0 \mid P(Z), D = 1) - \mathrm{E}(U_0 \mid P(Z), D = 0). \tag{7.14}$$

This is the "index sufficient" representation where $P(Z)$, or equivalently $H(Z)$, is the index.[54] An important question in the program evaluation literature is whether the selection bias can be characterized solely as a function of $P(Z)$ for different sets of Z, or if a more general conditioning set (Q,Z) is required to characterize this bias. In terms of the behavioral model of program participation and program outcomes presented in Section 6.2, the cost of participation, c, may play the role of V assuming that it is independent of other variables. Y_{0k} also could play that role provided that we condition on observed variables in forming the probability, and that the residual from this conditioning is independent of all the explanatory variables in the model.

Conventional econometric selection models (e.g., Amemiya, 1985) assume that the latent variables V, U_0, U_1 are symmetrically distributed around zero. The assumption of symmetry for U_0 and V implies that the bias $B(P(Z))$ is symmetric around $P(Z)$ equal to one-half. As shown by Fig. 8, in the normal selection model, if $P(Z)$ is symmetrically distributed around one-half, the average bias over symmetric intervals around that value is zero even though the pointwise bias is non-zero. If the values of $P(Z)$ for a sample of trainees were symmetrically distributed around one-half, the pointwise bias would be non-zero and the assumption justifying matching would not hold. Nonetheless, the selection bias would still average out to zero over any symmetric intervals of $P(Z)$ constructed around $P(Z) = 1/2$. Hence, the classical selection model justifies matching as a consistent estimator of (7.13) when it is defined over intervals of $P(Z)$ where the bias cancels out,

[54] See Heckman (1980) for the first derivation of this representation or Heckman and Robb (1985a, 1986a). Multiple decision rules for admission into a program require a multiple index model (Heckman and Robb, 1985a).

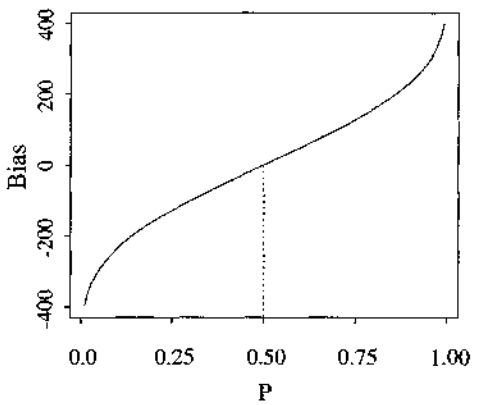

Fig. 8. Prototypical selection model, normal example: $B(P(X)) = \mathrm{E}(U_0 \mid P(X), D = 1) - \mathrm{E}(U_0 \mid P(X), D = 0)$. This is the index model where V and U_0 are assumed to be normal and $\sigma_V = 1$, $\sigma_{U_0} = 375$ and $\rho = \mathrm{cov}(U_0, V)/\sigma_{U_0} = 0.16$.

even though it would *not* justify matching for $\mathrm{E}(Y_1 - Y_0 \mid X, D = 1)$ defined pointwise for any points except those where the bias is zero.

To estimate the mean effect of treatment on the treated in the classic econometric selection model, we form the following regression based on Eq. (3.3):

$$\begin{aligned}\mathrm{E}(Y \mid Q, P(Z), D) &= \mathrm{E}(Y_1 D + Y_0(1 - D) \mid Q, P(Z), D) \\ &= g_0(Q) + D(g_1(Q) - g_0(Q)) + D(\mathrm{E}(U_1 \mid Q, P(Z), D = 1)) \\ &\quad + (1 - D)\mathrm{E}(U_0 \mid Q, P(Z), D = 0). \qquad (7.15)\end{aligned}$$

The conditional means of the error terms $\mathrm{E}(U_1 \mid Q, P(Z), D = 1)$ and $\mathrm{E}(U_0 \mid Q, P(Z), D = 0)$ are called *control functions* (Heckman and Robb, 1985a, 1986a). Under the assumptions that U_0, U_1 are statistically independent of Q and Z, these functions may be written as

$$\mathrm{E}(U_0 \mid Q, P(Z), D = 0) = K_0(P(Z)),$$

$$\mathrm{E}(U_1 \mid Q, P(Z), D = 1) = K_1(P(Z)).$$

Specific distributional assumptions about (U_0, V) and (U_1, V) produce specific functional forms for K_0 and K_1. Heckman and MaCurdy (1986) present a catalogue of parametric models including the normal sample selection model of Heckman (1976, 1979).

Under these conditions, Eq. (7.15) is really just two sample selection bias equations applied to non-participants and participants respectively:

$$\mathrm{E}(Y_0 \mid Q, P(Z), D = 0) = g_0(Q) + K_0(P(Z)), \qquad (7.16a)$$

$$\mathrm{E}(Y_1 \mid Q, P(Z), D = 1) = g_1(Q) + K_1(P(Z)). \qquad (7.16b)$$

The most common form of the model writes $g_0(Q)=Q\beta_0$ and $g_1(Q)=Q\beta_1$, but this is not strictly required. We can use the $D = 1$ and $D = 0$ samples to recover the parameters of the model.

Assuming that there is at least one exclusion restriction (a variable in Z not in Q), and that $K_0(P(Z))$ and $K_1(P(Z))$ are not perfectly collinear with Q, we can identify $g_0(Q)$ and $g_1(Q)$ up to intercepts for any K_0 and K_1 functions. The intercepts are not determined. Any intercept in $g_0(Q)$ can be allocated to K_0 and vice versa; the same remark applies to the allocation of intercepts between $g_1(Q)$ and K_1. To identify the intercepts, it is necessary to have some Z values, say Z_0, such that $K_0(P(Z_0)) = 0$ and some Z values, say Z_1, such that $K_1(P(Z_1)) = 0$. Using such values, one can identify the unique intercepts for g_0 and g_1, respectively (Heckman, 1990).[55] Another way to determine the intercepts is to assume specific functional forms for K_0 and K_1 that exclude intercept terms as in the conventional normal selection bias model.

Many non-parametric and semiparametric selection bias strategies have been proposed that do not impose functional form assumptions on K_0 and K_1. All of these strategies require that we identify the intercepts on sets Z_0 and Z_1, respectively. See the comprehensive surveys by Heckman (1990), Powell (1994), and Honoré and Kyriazidou (1997). Andrews and Schafgans (1998) extend a method proposed in Heckman (1990) to identify the intercepts.

With g_0 and g_1 in hand, we can estimate

$$\mathrm{E}(Y_1 - Y_0 \mid Q) = g_1(Q) - g_0(Q).$$

To form $\mathrm{E}(Y_1 - Y_0 \mid Q, P(Z), D = 1)$ observe that from the preceding analysis we know $g_0(Q)$, $g_1(Q)$ and

$$\mathrm{E}(U_1 \mid Q, P(Z), D = 1) = K_1(P(Z)).$$

We do not directly estimate $\mathrm{E}(U_0 \mid Q, P(Z), D = 1)$. However, under our assumptions about the (mean) independence of U_0 and (Q,Z), we can write

$$0 = \mathrm{E}(U_0 \mid Q, P(Z), D = 1)P(Z) + \mathrm{E}(U_0 \mid Q, P(Z), D = 0)(1 - P(Z)).$$

Because we know both the second term in this expression and $P(Z)$, we can form

$$\mathrm{E}(U_0 \mid Q, P(Z), D = 1) = -K_0(P(Z))\frac{1 - P(Z)}{P(Z)}.$$

Thus we can construct[56]

$$\mathrm{E}(Y_1 - Y_0 \mid Q, P(Z), D = 1) = g_1(Q) - g_0(Q) + K_1(P(Z)) + K_0(P(Z))\frac{1 - P(Z)}{P(Z)}.$$

[55] This type of identification on limit sets is sometimes called "identification at infinity" because for some models the values of Z_0 and Z_1 that set K_0 and K_1 to zero are $\pm\infty$.

[56] Björklund and Moffitt (1987) construct $\mathrm{E}(Y_1 - Y_0 \mid X, D = 1)$ in exactly this way for a normal selection model.

To estimate $E(Y_1 - Y_0 \mid Q, D = 1)$ we simply integrate out (average out) $P(Z)$ against the density of $P(Z)$ conditional on $D = 1$ and Q, which can be estimated. Thus, by making separability, exclusion and intercept identification assumptions, we can identify the parameter of interest (see Heckman et al. (1998b) for details.)

The control function method parameterizes the bias function $B(P(Z))$ in terms of $K_1(P)$ and $K_0(P)$ and estimates these functions along with the other parameters of the model. The dependence induced between U_0 and D operating through the V is called "selection on unobservables." The dependence induced between U_0 and D operating through dependence between Z and U_0 is termed "selection on observables" (Heckman and Robb, 1985a, 1986a). In this context, the method of matching assumes selection on observables, because conditioning on Z controls the dependence between D and U_0, producing a counterpart to (7.6a) for the residuals: $E(U_0 \mid Z, D = 1) = E(U_0 \mid Z, D = 0)$. When selection is on unobservables, it is impossible to condition on Z and eliminate the selection bias. We next turn to the method of instrumental variables which, like matching, assumes that selection only occurs on the observables.

7.4.3. The method of instrumental variables

The method of instrumental variables (IV) applied to estimate $E(Y_1 - Y_0 \mid X, D = 1)$ is a variant of the method of matching. It augments the X variables in matching with instruments Z so that

$$E(U_1 - U_0 \mid X, Z, D = 1) = E(U_1 - U_0 \mid X, D = 1), \tag{7.17a}$$

$$E(U_0 \mid X, Z) = E(U_0 \mid X) \tag{7.17b}$$

and that

$$\Pr(D = 1 \mid X, Z) \tag{7.17c}$$

depends in a non-trivial way on both X and Z. In particular, there must be at least two values of Z, say Z' and Z'', such that for any X where we seek to identify the parameter of interest, $\Pr(D = 1 \mid X, Z') \neq \Pr(D = 1 \mid X, Z'')$. We assume that (X,Z) satisfies the non-causality condition (7.A.1) replacing X in that condition with (X,Z).

Condition (7.17a) rules out any dependence between $U_1 - U_0$ and Z given X and D. It is implied by the condition

$$\Pr(D = 1 \mid X, Z, U_1 - U_0) = \Pr(D = 1 \mid X, Z).$$

The second condition (7.17b) says that U_0 may depend on X but not on Z. This is not a standard IV condition but it is analogous to the balance of bias condition in matching. Applying these conditions, we use the law of iterated expectations to write

$$E(Y \mid X, Z') = g_0(X)$$

$$+[g_1(X) - g_0(X) + E(U_1 \mid X, D = 1) - E(U_0 \mid X, D = 1)]\Pr(D = 1 \mid X, Z') + E(U_0 \mid X).$$

We can express $E(Y \mid X, Z'')$ similarly for the same X, but a different $Z = Z''$. By subtracting the $E(Y \mid X, Z'')$ from $E(Y \mid X, Z')$, we can form the following expression:

$$\frac{E(Y \mid X, Z') - E(Y \mid X, Z'')}{\Pr(D = 1 \mid X, Z') - \Pr(D = 1 \mid X, Z'')} = g_1(X) - g_0(X) + E(U_1 - U_0 \mid X, D = 1)$$

$$= E(Y_1 - Y_0 \mid X, D = 1). \tag{7.18}$$

Condition (7.17a) ensures us that when we further condition on Z, it does not affect the conditioning of $U_1 - U_0$ on $D = 1$ and X. Condition (7.17c) assures us that the denominator of the expression is not zero.

Observe that if we assume that $E(U_0 \mid X) = 0$ and $E(U_1 \mid X) = 0$ (so $g_0(X) = \mu_0(X)$ and $g_1(x) = \mu_1(X)$),[57] and if we assume that

$$(U_0, U_1) \perp\!\!\!\perp D \mid X, Z', \tag{7.19}$$

then IV also identifies

$$E(Y_1 - Y_0 \mid X) = g_1(X) - g_0(X) = \mu_0(X) - \mu_1(X),$$

the effect of treatment on a randomly chosen person with characteristics X. Under these assumptions, matching and IV are now indistinguishable except that IV augments the original X variables by Z.[58]

If individuals select into the program on the basis of the gain in unobservables, $U_1 - U_0$, or on the basis of variables that are (stochastically) dependent on the gain in unobservables, the conditions required for IV estimators to consistently estimate $E(Y_1 - Y_0 \mid X, D = 1)$ are not satisfied (Heckman and Robb, 1985a, 1986a,b; Heckman, 1997) unless $U_1 = U_0$ or $U_1 - U_0$ is unknown or not acted on at the time program participation decisions are made. If the instrument Z is correlated with the gain in unobservables, and if individuals base their participation decisions at least in part on that gain, then the instrument is correlated with the error in the outcome equation. For the parameter of interest, treatment on the treated, failure of (7.17a) produces:

$$E(Y_1 - Y_0 \mid X, Z, D = 1) = (g_1(X) - g_0(X)) + E(U_1 - U_0 \mid X, Z, D = 1).$$

Because the instrument enters the second term on the right hand side, it is not a valid instrument. The outcome equation may be written as

$$Y = g_0(X) + DE(Y_1 - Y_0 \mid X, Z, D = 1)$$

$$+ \{U_0 + D[(U_1 - U_0) - E(U_1 - U_0 \mid X, Z, D = 1)]\}.$$

[57] If $E(U_0 \mid X) = 0$, then (7.17b) is the more familiar IV condition $E(U_0 \mid X, Z) = E(U_0 \mid X) = 0$.

[58] Observe that even if $E(U_0 \mid X) \neq 0$ and $E(U_1 \mid X) \neq 0$, under conditions (7.17a) to (7.17c), IV identifies $E(Y_1 - Y_0 \mid X) = g_1(X) - g_0(X) + E(U_1 - U_0 \mid X)$.

The term in braces is the unobservable when the parameter of interest is the impact of training on the trained. For Z to be a valid instrument, it must be mean independent of this error term. But if the gain in unobservables determines participation, then Z conditional on $D = 1$ is related to the gain and the expectation of the error term conditional on Z is certainly not equal to zero. The implication of this result is that when the response to training varies among individuals, and the parameter of interest is the impact of treatment on the treated, the method of instrumental variables requires a strong behavioral assumption about how persons make their decisions about program participation.

To make this point more concretely, consider an example in which program evaluators use the distance between a person's residence and the training center as an instrument. They assume that the distance to the training center affects outcomes only through the participation indicator in the earnings equation. The problem that arises in the heterogeneous response framework is that we would expect persons who live far away from the training center to participate in training only when their expected gain from training is relatively large – large enough to offset their higher cost of participation. By contrast, persons closer to the training center, who therefore face a lower cost of participation, will have smaller average expected gains from training. As a result, if an individual participates in training, their post-training earnings also depend on how far away they live from a training center. Therefore the instrument, distance, is correlated with the unobserved component of the gain from training for those who take training ($D = 1$) even if it is not for a random sample of persons in the population. Put differently, knowing how far trainees live from a training center tells us something about their expected earnings even conditional on their training status, which means that distance from the training center is not a valid instrument in this case.[59]

7.4.4. The instrumental variable estimator as a matching estimator

Heckman (1998c) shows how most evaluation estimators, including IV estimators, can be interpreted as matching estimators using the weighting framework of Eqs. (7.8) and (7.10). To see the basic idea, consider the simple random coefficient model

$$Y = \beta(X) + \alpha D + U.$$

We define β and α as functions of X where $\mathrm{E}(U \mid X, D) \neq 0$. Assume a valid instrument Z that satisfies conditions (7.17a)–(7.17c). Then

$$\mathrm{E}(Y \mid X, Z) = \beta(X) + \mathrm{E}(\alpha \mid X, D = 1)\mathrm{E}(D \mid X, Z) + \mathrm{E}(U \mid X, Z).$$

Now we can express the outcome equation as follows:

[59] Notice that this is an alternative interpretation that explains the "discount rate bias" recently discussed by Card (1995). Instrumenting by distance to a school or a training center may *raise* the estimated return to schooling or training if responses to schooling or training are heterogeneous and persons act on this heterogeneity in enrolling in schooling or training programs.

$$Y = \beta(X) + \mathrm{E}(\alpha \mid X, D = 1)\mathrm{E}(D \mid X, Z) + U$$

$$+[\alpha - \mathrm{E}(\alpha \mid X, D = 1)][\mathrm{E}(D \mid X, Z) + W] + \mathrm{E}(\alpha \mid X, D = 1)W,$$

where $D = \mathrm{E}(D \mid X, Z) + W$ and where, under our assumptions, the error terms have mean zero conditional on X and Z.[60] If we have a valid instrument, then $\mathrm{E}(U \mid X, Z) = \mathrm{E}(U \mid X)$ and $\mathrm{E}(\alpha \mid X, Z, D = 1) = \mathrm{E}(\alpha \mid X, D = 1)$. To identify $\mathrm{E}(\alpha \mid X, D = 1)$ we may form pairwise comparisons between person i and *anyone* else, provided that the matched partner for i, say i', has the same X but a different $Z = Z'$, where

$$\mathrm{E}(D \mid X, Z) \neq \mathrm{E}(D \mid X, Z').$$

If this condition is satisfied, we may match a suitable i' to form the pairwise estimate of the gains as follows:

$$\frac{Y_i - Y_{i'}}{\mathrm{E}(D_i \mid X, Z_i) - \mathrm{E}(D_{i'} \mid X, Z_{i'})}.$$

Therefore,

$$\mathrm{E}\left[\frac{Y_i - Y_{i'}}{\mathrm{E}(D_i \mid X, Z_i) - \mathrm{E}(D_{i'} \mid X, Z_{i'})}\right] = \mathrm{E}(\alpha \mid X, D = 1).$$

Accordingly, we can write our estimate of $\mathrm{E}(\alpha \mid X, D = 1)$ as a weighted average of contrasts:

$$\hat{\alpha} = \sum_{i,i'}\left[\frac{(Y_i - Y_{i'})}{\mathrm{E}(D_i \mid X, Z_i) - \mathrm{E}(D_{i'} \mid X, Z_{i'})}\right]W(i, i') \tag{7.20}$$

for i,i' such that $\mathrm{E}(D_i \mid X, Z_i) \neq \mathrm{E}(D_i' \mid X, Z_i')$, and where the weights are given by

$$W(i, i') = \frac{(\mathrm{E}(D_i \mid X, Z_i) - \mathrm{E}(D_{i'} \mid X, Z_{i'}))^2}{\sum_{i,i'}(\mathrm{E}(D_i \mid X, Z_i) - \mathrm{E}(D_i \mid X, Z_{i'}))^2}.$$

Formally, we set

$$\frac{Y_i - Y_{i'}}{\mathrm{E}(D_i \mid X, Z_i) - \mathrm{E}(D_{i'} \mid X, Z_{i'})} = 0$$

for i,i', where $\mathrm{E}(D_i|X, Z_i) = \mathrm{E}(D_{i'}|X, Z_{i'})$ and we get the same result summed over all i,i' since for these cases $W(i, i') = 0$.

Eq. (7.20) reveals that propensity score matching with Z as the propensity score estimates $\mathrm{E}(\alpha \mid X, D = 1)$ by taking a weighted average of all i,i' contrasts for values of (X,Z)

[60] As we have stressed, all we need is that the error terms depend only on X in order to recover $\mathrm{E}(\alpha(X) \mid X, D = 1)$.

with distinct probability values. Instrumental variable estimation is just a weighted average of contrasts of conditional means constructed in terms of propensity scores. Observe that this method only requires (7.17b) and not that $E(U \mid X, Z) = 0$. Thus, like matching and randomized trials, the IV method does not eliminate conventional econometric exogeneity bias – it just balances the bias.

7.4.5. IV estimators and the local average treatment effect

Imbens and Angrist (1994) reinterpret the output of IV Eq. (7.18) as the effect of treatment on those who change state in response to a change in Z. It is a discrete approximation to the marginal treatment effect (3.14) previously discussed in Section 3.4 and defined as the effect of a marginal change of a policy on those induced to change state as a consequence of the policy. Keeping the conditioning on X implicit, their parameter is $E(Y_1 - Y_0 \mid D(z) = 1, D(z') = 0)$ where $D(z)$ is the conditional random variable D given $Z = z$, and where z' is distinct from z, so $z \neq z'$. This conditions on people who switch from "0" to "1" as a consequence of the change in Z. This parameter is termed "LATE" for Local Average Treatment Effect.

The LATE parameter has several non-standard features. It is *defined* by variation in an instrumental variable that is external to the outcome equation. Unlike the instrumental variables discussed in the preceding section, in LATE, different instruments *define* different parameters. In the traditional IV literature, Z is used to identify the effect of X on outcomes. In LATE, variation in Z defines the parameter and no distinction between X and Z is made. When the instruments are indicator variables that denote different policy regimes, or when the instruments are different levels of intensity of a policy within a given regime (i.e., the level of φ in terms of the analysis of Section 3.4), LATE identifies the response to policy changes for those who change their program participation status in response to the policy change. When the instruments refer to personal or neighborhood characteristics used to predict an endogenous variable, say schooling in an earnings equation, LATE has a less clear cut interpretation and its relevance for policy analysis is questionable.

The measured variation in Z among people could be due to their choices of Z. If distance to the nearest school or training center is the instrument, LATE estimates the effect of variation in the distance to school on the earnings gain of persons who are induced to change their schooling or training status as a consequence of the different commuting costs they face. If a personal characteristic is used as an instrument, for example, family income, the parameter defines the marginal change in the outcome with respect to the variation in family income among those who would have changed their state in response to the sample variation in family income.

To define the LATE parameter more precisely, let $D(z)$ be the conditional random variable D given $Z = z$. (Recall that conditioning on X is kept implicit in this section). Since $D(z)$ is defined conditional on a particular realization of $Z = z$, it is independent of Z.[61] Imbens and Angrist (1994) assume that:

$(Y_0, Y_1, D(z))$ are independent Z and Pr(D=1|Z=z) is a non-trivial function of Z where these random variables are understood to be defined conditional on X. (7.IA.1)

As a consequence of this assumption, for a given person (with fixed Y_1, Y_0), and recalling that for $Z = z$, $Y = Y_0(1 - D(z)) + Y_1 D(z)$, we may write

$$\mathrm{E}(Y \mid Z = z) - \mathrm{E}(Y \mid Z = z') = \mathrm{E}[D(z)Y_1 + (1 - D(z))Y_0 \mid Z = z]$$

$$-\mathrm{E}[D(z')Y_1 + (1 - D(z'))Y_0 \mid Z = z']$$

$$= \mathrm{E}((D(z) - D(z'))(Y_1 - Y_0)). \tag{7.21}$$

The final step follows from assumption (7.IA.1) and depends crucially on the conditional independence of Y_1, Y_0 and $D(z)$ from Z.

In the Imbens–Angrist thought experiment, all of the random variables in the expression are defined for the same person. Thus for different values of $Z = z$, Y_1 and Y_0 do not change and $\{D(Z)\}$ for z in the support of Z is a collection of not necessarily independent random variables produced by changing Z and either not changing any other random variable or changing them only in the way specified in assumption (7.IA.2) below. In terms of the index model of discrete choice theory with index function $H(Z,V)$, which may be a net profit or net utility function, we have

$$D = \mathbb{1}(H(z, V) \geq 0) \tag{7.22}$$

and V is a random variable. In the Imbens and Angrist (1994) thought experiment, V stays fixed while z is varied.

From Eq. (7.21) it follows that

$$\mathrm{E}(Y \mid Z = z) - \mathrm{E}(Y \mid Z = z') = \mathrm{E}(Y_1 - Y_0 \mid D(z) - D(z') = 1)\Pr(D(z) - D(z') = 1)$$

$$+\mathrm{E}(Y_1 - Y_0 \mid D(z) - D(z') = -1)\Pr(D(z) - D(z') = -1). \tag{7.23}$$

This is the total effect on the outcome measure of a change in Z, including the effect on those induced to enter the program and the effect on those induced to leave the program. In terms of our discussion in Section 3.4, if Z is a policy variable, this produces the net effect of a change in Z on the aggregate measure of Y. This is one of the necessary ingredients for a cost benefit analysis of the effect of a marginal change in a policy variable on outcomes.

Imbens and Angrist (1994) break up the total effect into two terms: $\mathrm{E}(Y_1 - Y_0 \mid D(z) - D(z') = 1)$ and $\mathrm{E}(Y_1 - Y_0 \mid D(z) - D(z') = -1)$, defined for those induced into the program and induced out of it, respectively, and they present conditions that make it

[61] For two random variables (J,K) let f be the density (or frequency). Then $f(J,K) = f(J \mid K)f(K)$ so J given K is statistically independent of K although $f(J \mid K)$ may be functionally dependent on K.

possible to identify one of these. To identify the Imbens and Angrist (1994) "causal" parameters, a second assumption about the hypothetical random variables is required:

For all z, z' in the support of z, either $D(z) \geq D(z')$ for all persons or
$D(z) \leq D(z')$ for all persons. (7.IA.2)

Assuming that the denominator is not zero, this monotonicity assumption zeros out one of the two terms in (7.23). The assumption regarding the denominator is a technical condition. Even if the denominator is zero the program may have an effect on the aggregate through a shift in the composition of participants and non-participants. The variation across z and z' is made holding the error term constant. Condition (7.IA.2) makes either $\Pr(D(z) - D(z') = 1)$ or $\Pr(D(z) - D(z') = -1)$ zero for everyone. Thus, under their conditions the effect of a change in Z is to shift people into one sector or the other but not both. Suppose $D(z) \geq D(z')$, then $\Pr(D(z) - D(z') = -1) = 0$ and using (7.23) we obtain

$$\mathrm{E}(Y_1 - Y_0 \mid D(z) - D(z') = 1) = \frac{\mathrm{E}(Y \mid Z = z) - \mathrm{E}(Y \mid Z = z')}{\Pr(D = 1 \mid Z = z) - \Pr(D = 1 \mid Z = z')}. \tag{7.24}$$

If the monotonicity assumption is violated, IV estimates a weighted average of the LATE arising from people flowing from 0 to 1 and a reverse LATE arising from people flowing from 1 to 0, with the weights being

$$\frac{\Pr(D(z) - D(z') = 1)}{\Pr(D = 1 \mid Z = z) - \Pr(D = 1 \mid Z = z')} \quad \text{and} \quad \frac{\Pr(D(z) - D(z') = -1)}{\Pr(D = 1 \mid Z = z) - \Pr(D = 1 \mid Z = z')},$$

respectively. Because LATE is *defined* in terms of population moments, it can be consistently estimated by instrumental variables methods replacing population moments by sample moments.

Comparing (7.18) with (7.24) reveals that "LATE" looks like what the standard IV converges to except for one important difference: the LATE parameter is z dependent. Both the LATE and $\mathrm{E}(Y_1 - Y_0 \mid X, D = 1)$ are identified by taking the ratio of the change in the outcome induced by Z and dividing by the change in the probability of being in sector 1 induced by $Z = z$. The parameter $\mathrm{E}(Y_1 - Y_0 \mid X, D = 1)$ does not depend on Z while the LATE parameter does. Observe further that if conditions (7.17a) through (7.17c) are satisfied, the LATE estimator also identifies $\mathrm{E}(Y_1 - Y_0 \mid X, D = 1)$. Thus, in the case of a common coefficient model, or in the case where responses to training are heterogeneous, but not acted on by agents, LATE identifies $\mathrm{E}(Y_1 - Y_0 \mid X, D = 1) = \mathrm{E}(Y_1 - Y_0 \mid X)$.

Condition (7.IA.2) is satisfied if (7.22) characterizes choices. It is also satisfied by any index $IN = H(z, V_z)$ where

$$D(z) = 1(IN > 0 \mid Z = z)$$

characterizes participation in the program being evaluated, provided that H is increasing in

z, V_z is increasing in z and H is increasing in V_z. This would be satisfied in the case of a scalar z if

$$V_z = V_{z'} + \sigma(z),$$

for $z > z'$, where $\sigma(z)$ is a random variable with $\sigma(z) > 0$ when $z > z'$. If, however, $\sigma(z)$ is permitted to be both positive and negative, condition (7.IA.2) would not be satisfied.

The Roy model estimated by Heckman and Sedlacek (1985) has a decision rule of the form (7.22) or (6.5):

$$IN = Y_1 - Y_0 + k(z)$$

and

$$D = 1(IN > 0 \mid Z = z).$$

If $k(z)$ is monotonic in z, this decision rule produces a model consistent with (7.IA.2). To see this, assume that Y_1 and Y_0 are continuous random variables and that Z is independent of $(Y_1 - Y_0)$ so that the conditions of (7.IA.1) are satisfied. In the Imbens and Angrist (1994) thought experiment that defines their estimator, $Y_1 - Y_0 = V$ is fixed and different realizations of Z are considered. In this set up, the event $D(z) - D(z') = 1$ is described by the inequalities

$$Y_1 - Y_0 + k(z) > 0 \qquad \text{and} \qquad Y_1 - Y_0 + k(z') < 0$$

so that

$$-k(z') > Y_1 - Y_0 > -k(z)$$

and the condition $D(z) - D(z') = 1$ induces a partition of $Y_1 - Y_0$. Now the LATE "causal parameter" is

$$\mathrm{E}(Y_1 - Y_0 \mid D(z) - D(z') = 1) = \mathrm{E}(Y_1 - Y_0 \mid -k(z) < Y_1 - Y_0 < -k(z')),$$

which clearly depends on the choice of z and z'.[62] This example is a clear illustration of how under its assumptions LATE sidesteps the problem that Z is not a valid instrument for the treatment on the treated parameter in the Roy model. It estimates a different parameter that under its assumption approximates part of the marginal effect of a policy change derived in Section 3.4.

Consider once more our example of distance from the training center as an instrument for estimating the impact of training on earnings. If the LATE assumptions apply, but assumption (7.17a) does not, then LATE identifies the impact of commuting distance variation on training outcomes for those induced to participate by the change in the commuting distance. The LATE estimator with distance to the training center used as Z

[62] This example illustrates the point that statistical independence of two random variables does not imply their functional independence.

does not identify the impact of training for other samples or the LATE associated with different instruments.[63]

In general, the LATE parameter depends on the particular choice of the z and z' as well as X. Factors external to the outcome equation define the LATE parameter and a different parameter is produced for each choice of z and z'. If there are multiple instruments, there are multiple parameters. Additional instruments do not improve the efficiency with which a fixed parameter is estimated as they would in standard "policy invariant" structural models. Instead different instruments define different parameters. However, we have presented cases where the instruments are policy changes and LATE identifies a policy relevant parameter.

Heckman and Vytlacil (1999a,b) introduce a new parameter – the Local Instrumental Variable (LIV) parameter – which is a limit form of LATE when the instruments are continuous. A variety of evaluation parameters including LATE can be generated from it by using different weighting schemes. LIV is the fundamental building block of evaluation analysis. Heckman and Vytlacil (1999a) use LIV to bound parameters when treatment effects are not identified.

Imbens and Angrist (1994) claim that their identifying assumptions are much weaker than the more familiar identifying assumptions used in econometrics based on index models or latent variables crossing thresholds. In fact, their assumptions are *equivalent* to assuming a latent variable so there is no added generality in their approach (see Vytlacil, 1999).

7.4.6. Regression discontinuity estimators

Regression discontinuity estimators constitute a special case of "selection on observables." Originally introduced by Campbell and Stanley (1963), evaluations based on them have been presented by Goldberger (1972), Cain (1975), Barnow et al. (1980), Trochim (1984) and, more recently, by van der Klaauw (1997) and Hahn et al. (1998). In this model, treatment depends on some observed variable, S, according to a known, deterministic rule, such as $D = 1$ if $S < \bar{S}$, $D = 0$ otherwise. If Y_0 depends on S, and if $\alpha \neq 0$, then this assignment rule will induce a discontinuity in the relationship between $Y = Y_0 + D\alpha$ and S at the point $S = \bar{S}$.[64] Two features distinguish this case from the standard selection on observables case discussed in Section 7.4.1. First, there is no common support for participants and non-participants. For all values of S, $\Pr(D = 1 \mid S) \in \{0,1\}$. Thus, matching is impossible. Recall our example in Section 6.3 where the analyst knows (c, Y_{0k}). Conditioning on both of these variables violates the assumption (7.5). Thus the regression discontinuity estimator takes over when there is selection on observables but the overlapping support condition required for matching breaks down. Alternatively, the regression

[63] If the same parameter is estimated for all choices of commuting distances, then (7.17a) holds and the LATE estimator, which is formally equivalent to the IV estimator, recovers the impact of treatment on the treated. This is the basis for a test of whether the LATE is equivalent to the impact of treatment on the treated over the range of distance values for which it is estimated.

[64] We consider the assignment rule $S < \bar{S}$ for simplicity. The case with $S > \bar{S}$ is symmetric.

discontinuity design estimator is a limit form of matching at one point. Second, the selection rule is assumed to be deterministic and known.

Barnow et al. (1980), present a simple example of this estimator. They consider a hypothetical enrichment program for disadvantaged children based loosely on the US Head Start program. Children with family incomes below a cutoff level receive the program, all other children do not. The outcome variable of interest is the children's test scores. As shown in Fig. 9, the underlying relationship between test scores and family income is assumed to be linear. The line segment above the cutoff level reflects this relationship, which would continue (as shown by the broken line) to lower levels of family income in the absence of the program. The discontinuity in the regression line at the cutoff point represents the effect of the program, which is assumed to be a constant α. Under the assumptions of a common effect model and of linearity in the relationship between children's test scores and pre-program family income, α can be estimated without bias by OLS estimation of:

$$Y = \beta_0 + \alpha D + \beta_1 S + U. \tag{7.25}$$

Now consider the random coefficient case where α varies. We let α_i be the value of α for person i. The deterministic selection rule assumed in the regression discontinuity design precludes individuals choosing to participate in the program based on α_i. However, if α_i varies with S, then the mean impact of the treatment on the treated may differ from the mean impact of the treatment on a randomly selected person in the population. Due to the lack of a common support, no information about the impact of treatment on the untreated is available except at the point of discontinuity other than through extrapolation of the impact estimated for participants via functional form assumptions. Such an extrapolation

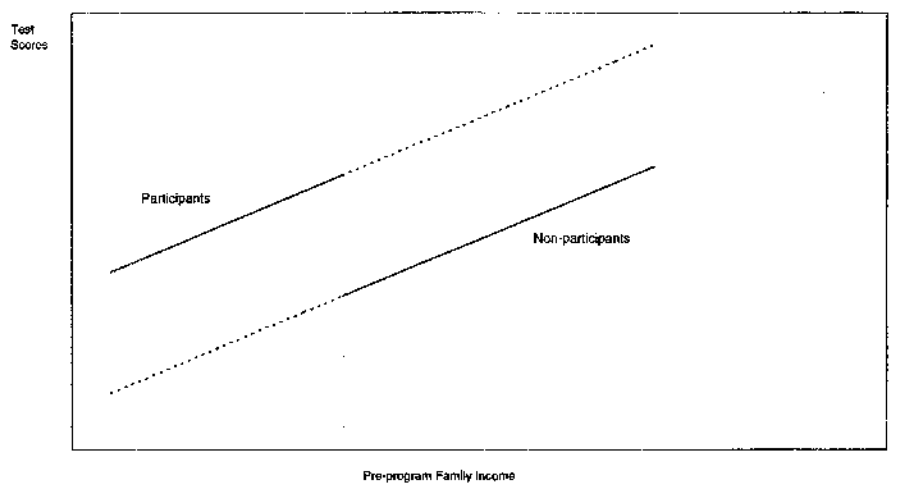

Fig. 9. Barnow et al. (1980): Head Start impact extrapolation example.

is illustrated in Fig. 9 in the upper broken line. This is a limitation of the estimator because one policy change of interest would be to increase the cutoff level to allow persons presently excluded at the margin to be included in the program. Some continuity has to be assumed to use the estimator in this situation. If it is assumed that the functional form of the relationship $Y_0(S)$ is known or can be determined using the available data, then we can estimate the impact of the treatment as a function of S, $\alpha(S)$, for persons in the program as the difference between the extrapolated $Y_0(S)$ and the observed outcomes of participants at each value of S. In the simplest case, if α_i is a linear function of S, then OLS estimation of

$$Y = \beta_0 + \alpha_0 D + \beta_1 S + \alpha_1 DS + U \tag{7.26}$$

will yield unbiased estimates of the linear relationship $\alpha_i = \alpha_0 + \alpha_1 S$ for participants. With knowledge of this relationship, we can readily determine the effects of a policy of cutting back the program by changing the cutoff point to exclude more people.

The most important issue in the application of the regression discontinuity design estimator is the extent to which the functional form of $Y_0(S)$ is known (possibly from samples not subject to treatment) or can be estimated. The older literature (e.g., Cook and Campbell, 1979) considers various methods such as selecting among polynomials in S through a combination of formal testing and visual inspection of the data. The more recent literature (e.g., Hahn et al., 1998) avoids this problem by estimating the impact of the treatment locally at the cutoff point using non-parametric methods. The former approach has the advantage of putting all of the data to use in identifying $Y_0(S)$; conditional on choosing the correct functional form, the parametric approach yields more precise estimates. The non-parametric approach has the advantage of avoiding extrapolation bias. In the random coefficient case, the non-parametric approach obtains only a local average treatment effect – the effect of treatment on participants with values of S close to $\bar{S}$. With multiple points of discontinuity, however, this problem becomes less severe. The parametric approach can still identify $\alpha(S)$ among participants as the difference between the observed relationship between Y and S conditional on $D = 1$ and the relationship $Y_0(S)$ estimated using the non-participants.

Another issue that arises in the analysis of regression discontinuity designs are so-called "fuzzy" assignment rules where assignment to the treatment is not a completely deterministic function of S. Except in the case of random variation in assignment conditional on S, fuzziness in the assignment rule changes this problem from one of selection on observables to one of selection on unobservables, conditional on S. The general methods for dealing with selection on unobservables discussed in this chapter can be applied in this situation, but much of the simplicity of the regression discontinuity design is lost (see the discussion in van der Klaauw, 1997). Still, the discontinuity assumed for S can aid in identifying parameters. For a random effect model, and under local monotonicity, Hahn et al. (1998) identify a LATE effect.

A final issue that arises in regression discontinuity designs concerns non-participation in the program by persons whose values of S make them eligible for it. Unless the common effect model is assumed, or the random effect model is assumed but participation does not

depend on the person-specific component of the impact, the simple estimation schemes described above no longer identify the mean impact of the treatment on persons satisfying the cutoff condition, even if the functional form of $Y_0(S)$ is known. They do, however, provide unbiased estimates of the impact of the treatment on the treated. If we seek to identify the impact of treatment on all persons below the cutoff point, $S < \bar{S}$, which would be of interest in regard to proposals to increase participation among persons already eligible (i.e., raising the "take-up rate"), we must apply modified versions of the non-experimental methods discussed in this section.

7.5. Using aggregate time series data on cohorts of participants to evaluate programs

For the model of Section 6.3 with one possible program enrollment period over the life-cycle, (e.g., schooling, or army service), and for many other models, it is sometimes possible to identify the effect of treatment on the treated using only data on cohort means, without knowing the treatment status of any individual in the cohort. As noted in Section 3.4, in principle one can evaluate a program using aggregate time series data and thereby avoid the selection problem. Initially, assume a time homogeneous environment. Estimates of the aggregate cohort mean outcomes formed in two or more cross-sections of unrelated persons measured before and after the age where participation in the program is possible can be used to obtain estimates of the effect of treatment on the treated free of selection bias even if the training status of each person is unknown so long as the cohort proportion of trainees is known or can be consistently estimated. With more data, the time homogeneity assumption can be partially relaxed, as we will demonstrate.

Assuming a time homogeneous environment and access to repeated cross-section data governed by random sampling, it is possible to identify $\alpha = \mathrm{E}(Y_1 - Y_0 \mid D = 1)$ (a) without any instrumental variables, (b) without need to specify the joint distribution of U_0, U_1 and V, and (c) without any need to know which individuals in the sample enrolled in training. However, the proportion of training must be known or consistently estimable (Heckman and Robb, 1985a, 1986b.)

To show how this is possible, suppose that no regressors appear in the earnings function.[65] Assuming that random sampling generates the data, the expectation of the cohort means (denoted by "-") is

$$\mathrm{E}(\bar{Y}_t) = \mathrm{E}(\beta_t + \alpha D + U_t) = \beta_t + \mathrm{E}(\alpha \mid D = 1)P, \qquad \text{for } t > k,$$

and

$$\mathrm{E}(\bar{Y}_{t'}) = \mathrm{E}(\beta_{t'} + U_{t'}) = \beta_{t'}, \qquad \text{for } t' < k,$$

[65] If regressors appear in the earnings functions, condition on X. See Heckman and Robb (1985a, 1986a) for the general case.

where $P = \Pr(D = 1)$. In a time homogeneous environment, $\beta_t = \beta_{t'}$ and

$$\frac{\mathrm{E}(\bar{Y}_t) - \mathrm{E}(\bar{Y}_{t'})}{P} = \mathrm{E}(\alpha \mid D = 1).$$

Replacing sample means with population means defines the estimator. The estimator can be formed within X strata. This is a grouping estimator that averages out the error term. Nowhere does it exploit any covariance term to identify the parameter. Hence, it is possible to identify the parameter when U is correlated with D and there is no conventional instrumental variable.

With more than two years of repeated cross-section data, one can apply the same principles to identify $\mathrm{E}(\alpha \mid D = 1)$ while relaxing the time homogeneity assumption. For instance, suppose that the time trend for cohort mean earnings lies on a polynomial of order $L - 2$:

$$\beta_t = \pi_0 + \pi_1 t + \cdots + \pi_{L-2} t^{L-2}.$$

From L temporally distinct cross-sections, it is possible to consistently estimate the $L - 1$ π-parameters and $\mathrm{E}(\alpha \mid D = 1)$ provided that the number of observations in each cross-section becomes large and there is at least one pre-program and one post-program cross-section.

If the effect of training differs across periods, it is still possible to identify $\mathrm{E}(\alpha_t \mid D = 1)$, provided that the environment changes in a "sufficiently regular" way. For example, suppose

$$\beta_t = \pi_0 + \pi_1 t,$$

$$\mathrm{E}(\alpha_t \mid D = 1) = \phi_0(\phi_1)^{t-k}, \qquad t > k.$$

In this case, π_0, π_1, ϕ_0, and ϕ_1 are identified from the means of four cross-sections, as long as at least two of these means come from a pre-program period and two come from successive post-program periods.[66]

Heckman and Robb (1985a, 1986b) state the conditions required to consistently estimate $\mathrm{E}(\alpha_t \mid D = 1)$ using repeated cross-section data on cohort aggregates which do not record the training identity of individuals under general conditions about cohort and time effects. Section 7.7 studies the sensitivity of this class of estimators to violations of the random sampling assumption.

7.6. Panel data estimators

Access to repeated observations on the same persons followed over time enables analysts to exploit the time series properties of the outcome equations and their relationship with

[66] Heckman and Robb (1985a) show how to solve the four equations for means in terms of the four unknown parameters.

program participation equations. Like the classical econometric selection bias estimators, panel data estimators exploit additive separability between model and error.

This subsection consists of four parts. In the first part, we consider panel data estimators for the common coefficient model (3.10). We allow α to depend on X but we assume only one error term $U_{1t} = U_{0t} = U_t$. A model with two errors in (3.10), $U_t = DU_{1t} + (1 - D)U_{0t}$, complicates the analysis and alters the conclusions reached for the simpler case of a single error term. This case requires a separate analysis because many longitudinal estimators are not robust to the introduction of two regime-specific error terms into the model.

In the second part, we extend the panel data models to apply to repeated cross-section data. We demonstrate how many conventional panel data evaluation estimators can be applied to repeated cross-sections of the same populations sampled over time. This is fortunate since repeated cross-section data are much more commonly available around the world than are panel data. In the third part, we extend these results to allow for a two component model, so that there is heterogeneity in responses to program participation on unmeasured outcomes ($U_{0t} \neq U_{1t}$). Finally, in the fourth part we show how the panel data estimators can be placed within the matching framework of Section 7.4.1.

7.6.1. Analysis of the common coefficient model

We start the analysis by assuming model (3.10) with $U_{1t} = U_{0t} = U_t$ so that $Y_{1t} - Y_{0t} = \alpha$ but α may depend on X, $\alpha(X)$.We consider more general cases below. The cases considered in this section are the familiar models used in conventional panel data analysis.

7.6.2. The fixed effects method

We begin our analysis with the conventional fixed effect model. Eq. (6.7) presents the key identifying assumption of the fixed effect method. If we allow Eq. (3.10) to include observed characteristics, the identifying assumption is:

$$\mathrm{E}(U_{0t} \mid X, D = 1) = \mathrm{E}(U_{0t'} \mid X, D = 1), \qquad \text{for some } t > k > t'. \tag{6.8'}$$

Recall that k is the period of program participation. Suppose that this condition holds and the analyst has access to 1 year of pre-program and 1 year of post-program outcome data. Regressing the difference between the outcomes in the years t and t' on a dummy variable for training status produces a consistent estimator of α. (This method is well exposited in Hsiao, 1986.) A variety of efficient estimators have been developed that exploit the multiplicity of contrasts that are sometimes available.

Some program participation rules and error processes for earnings justify condition (6.8'). For example, consider a certainty environment in which the earnings residual has a permanent-transitory structure:

$$U_t = \phi + \varepsilon_t,$$

where ε_t is a mean zero random variable independent of all other values of $\varepsilon_{t'}$ for $t \neq t'$, and is distributed independently of ϕ, a mean zero person-specific time-invariant random

variable. Assuming that the V ($=c + Y_{0k}$) in participation rule (6.4) are distributed independently of all ε_t, except possibly for ε_k, condition (6.8′) will be satisfied provided that decision rule (6.3) generates participation. However, this condition is violated if there are imperfect credit markets as in Section 6.3.4. With two periods of data (in t and t', $t > k > t'$), α is identified. With more periods of panel data, the model is overidentified and hence we can test condition (6.8′). See the discussion in Hsiao (1986).

The permanent-transitory error structure is very special. As already discussed in Section 6, much evidence speaks against this error specification as a description of earnings residuals. (See also the discussion of the evidence in Section 8.4.) This method is crucially dependent on additivity of the errors, strong assumptions about program participation rules and special assumptions about the time series properties of the errors. Thus it is not surprising that LaLonde (1986) finds the method to be one of the least reliable non-experimental estimators for evaluating training programs.

7.6.3. U_t follows a first-order autoregressive process

We consider a more general model and assume that U_t follows the first-order autoregression given by Eqs. (7.1a)–(7.1c). Substitution into (3.10) yields

$$Y_t = [X_t - (X_{t'}\rho^{t-t'})]\beta + (1 - \rho^{t-t'})D\alpha + \rho^{t-t'}Y_{t'} + \left\{\sum_{j=0}^{t-(t'+1)} \rho^j \varepsilon_j\right\}, \qquad \text{for } t > t' > k. \tag{7.27}$$

This expression is an alternative form of (7.2) that includes regressors. Assume further that either (i) the perfect foresight rule of Eq. (6.3) determines enrollment and the ε_j are distributed independently of X or (ii) that the post-k ε_t are not known at k, and are forecast to have zero means. (Heckman and Wolpin (1976) invoke similar assumptions in their analysis of affirmative action programs.) If the X are independent of $\varepsilon_{j'}$ for all j,j',[67] then least squares applied to Eq. (7.27) consistently estimates α in large samples.[68] Unlike the fixed effects model, the autoregressive model does not require preprogram earnings and hence can be used to evaluate schooling or training programs for youth. As is the case with the fixed effect estimator, the model becomes overidentified (and hence testable) for panels with more than two time periods. If we assume imperfect credit markets of the form presented in Section 6.3.4, the estimator is inconsistent because participation depends on all lagged and future ε_t and D is correlated with the error in (7.27).

7.6.4. U_t is covariance stationary

The next procedure invokes an assumption about the time series properties of the error that is implicitly used in many papers on training (Ashenfelter, 1978; Bassi, 1983), and exploits

[67] This condition can be weakened to mean independence: $E(\varepsilon_j \mid X_1, \ldots, X_T) = 0$ for all j.

[68] A non-linear regression that imposes restrictions across coefficients increases efficiency.

the assumption in a novel way (Heckman and Robb 1985a, 1986a). We assume the following:

(a) U_t is covariance stationary so $E(U_t U_{t-j}) = \sigma_j$ for $j \geq 0$;

(b) there is access to at least two observations on pre-program earnings in t' and $t' - j$ as well as one observation on post-program earnings in t where $t - t' = j$; and

(c) $E(U_{t'} \mid D = 1)P \neq 0$, where $P = \Pr(D = 1)$.

Unlike the two previous models, here we make no assumptions about the appropriate participation rule or about the stochastic relationship between U_t and the cost of enrollment in (3.10) or (6.3). We can define this model conditional on X values.

We write the model as

$$Y_t = \beta_t + D\alpha + U_t, \qquad \text{for } t > k,$$

$$Y_{t'} = \beta_{t'} + U_{t'}, \qquad \text{for } t' < k,$$

where β_t and $\beta_{t'}$ are period-specific shifters and the conditioning on X is kept implicit.

Using a random sample of pre-program earnings from periods t' and $t' - j$, we can consistently estimate $\sigma_j = \mathrm{Cov}(Y_{t'}, Y_{t'-j})$ using the least squares residuals. If $t > k$ and $t - t' = j$, so that the post-program earnings data are as far removed in time from t' as t' is removed from $t' - j$, the covariance $\mathrm{Cov}(Y_t, Y_{t'})$ satisfies

$$\mathrm{Cov}(Y_t, Y_{t'}) = \sigma_j + \alpha P E(U_{t'} \mid D = 1), \qquad \text{for } t > k > t'.$$

The covariance between D and $Y_{t'}$ is

$$\mathrm{Cov}(Y_{t'}, D) = PE(U_{t'} \mid D = 1), \qquad \text{for } t' < k.$$

Assuming $E(U_{t'} \mid D = 1)P \neq 0$ for $t' < k$, we obtain

$$\alpha = \frac{\mathrm{Cov}(Y_t, Y_{t'}) - \mathrm{Cov}(Y_{t'}, Y_{t'-j})}{\mathrm{Cov}(Y_{t'}, D)}.$$

Using sample moments in place of population moments defines the estimator. For panels of sufficient length (e.g., more than two pre-program observations or more than two post-program observations), the stationarity assumption can be tested. Increasing the length of the panel converts a just-identified model to an overidentified one. Heckman and Robb (1985a) consider a variety of other assumptions that exploit the time series properties of the panel data including factor structure models for error processes.

7.6.5. Repeated cross-section analogs of longitudinal procedures

We can apply most longitudinal procedures to repeated cross-section data. Such data are cheaper to collect, and they do not suffer from the problems of non-random attrition which often plagues panel data.[69] The previous section presented longitudinal estimators of α

[69] These points were first made in Heckman and Robb (1985a, 1986a).

that are based on identifying moment conditions. In all cases but one, however, we can identify α with repeated cross-section data. Heckman and Robb (1985a, 1986b) give many additional examples of longitudinal estimators which can be implemented on repeated cross-section data.

7.6.6. The fixed effect model

As in Section (7.6.2), assume that condition (6.8′) holds so that

$$\mathrm{E}(U_t \mid D = 1) = \mathrm{E}(U_{t'} \mid D = 1),$$

$$\mathrm{E}(U_t \mid D = 0) = \mathrm{E}(U_{t'} \mid D = 0),$$

for all t,t' such that $t > k > t'$. As before, we can condition on X. $\mathrm{E}(Y_t \mid D = 1)$ is the mean outcome of participants in year t and $\mathrm{E}(Y_t \mid D = 0)$ is the mean outcome of non-participants in year t, with sample counterparts $\bar{Y}_t^1$ and $\bar{Y}_t^0$ respectively. The parameter can be written in terms of population moments as

$$\alpha = [\mathrm{E}(Y_t \mid D = 1) - \mathrm{E}(Y_t \mid D = 0)] - [\mathrm{E}(Y_{t'} \mid D = 1) - \mathrm{E}(Y_{t'} \mid D = 0)]$$

with sample counterpart

$$\hat{\alpha} = (\bar{Y}_t^{(1)} - \bar{Y}_t^{(0)}) - (\bar{Y}_{t'}^{(1)} - \bar{Y}_{t'}^{(0)}).$$

Assuming random sampling, consistency of $\hat{\alpha}$ follows immediately. As in the case of the longitudinal version of this estimator, with more than two cross-sections, condition (6.8′) can be tested.

In one respect this example is contrived. It assumes that in the pre-program cross-sections we know the identity of future trainees. Such data might exist (e.g., individual person records can be matched to subsequent training records). One advantage of longitudinal data for identifying and estimating α is that we know the training status of all persons without resort to further sampling or matching of records across different data sources.

7.6.7. The error process follows a first-order autoregression

Suppose, instead, that U_t follows a first-order autoregressive process given by Eq. (7.1c) and that

$$\mathrm{E}(\varepsilon_t \mid D) = 0, \qquad \text{for } t > k.$$

It is possible to identify α with three successive post-program cross-sections in which the identity of trainees is known.

To establish this result, let the three post-program periods be t, $t + 1$ and $t + 2$. Assuming, as before, that no regressor appears in Eq. (7.2), or, alternatively, conditioning on X, we obtain:

$$\mathrm{E}(Y_j \mid D = 1) = \beta_j + \alpha + \mathrm{E}(U_j \mid D = 1),$$

$$\mathrm{E}(Y_j \mid D = 0) = \beta_j + \mathrm{E}(U_j \mid D = 0),$$

From condition (7.1c),

$$\mathrm{E}(U_{t+1} \mid D = 1) = \rho \mathrm{E}(U_t \mid D = 1),$$

$$\mathrm{E}(U_{t+1} \mid D = 0) = \rho \mathrm{E}(U_t \mid D = 0),$$

$$\mathrm{E}(U_{t+2} \mid D = 1) = \rho^2 \mathrm{E}(U_t \mid D = 1),$$

$$\mathrm{E}(U_{t+2} \mid D = 0) = \rho^2 \mathrm{E}(U_t \mid D = 0),$$

Then

$$\rho = \frac{[\mathrm{E}(Y_{t+2} \mid D = 1) - \mathrm{E}(Y_{t+2} \mid D = 0)] - [\mathrm{E}(Y_{t+1} \mid D = 1) - \mathrm{E}(Y_{t+1} \mid D = 0)]}{[\mathrm{E}(Y_{t+1} \mid D = 1) - \mathrm{E}(Y_t \mid D = 0)] - [\mathrm{E}(Y_t \mid D = 1) - \mathrm{E}(Y_t \mid D = 0)]} \tag{7.28a}$$

and

$$\alpha = \frac{[\mathrm{E}(Y_{t+2} \mid D = 1) - \mathrm{E}(Y_{t+2} \mid D = 0)] - \rho[\mathrm{E}(Y_{t+1} \mid D = 1) - \mathrm{E}(Y_{t+1} \mid D = 0)]}{1 - \rho}. \tag{7.28b}$$

Replacing population moments by sample moments defines the estimator.[70]

For this model, the advantage of longitudinal data is clear. Only two time periods of longitudinal data are required to identify α, but three periods of repeated cross-section data are required to estimate the same parameter. However, if Y_t is subject to measurement error, the apparent advantages of longitudinal data become less clear. Repeated cross-section estimators are robust to mean zero measurement error in the variables. The longitudinal regression estimator discussed in the preceding section does not identify α unless the analyst observes earnings without error or has access to instruments to adjust for the measurement error. Given 3 years of longitudinal data and assuming that measurement error is serially uncorrelated, one could instrument Y_t, in Eq. (7.2), using earnings in the earliest year as an instrument. This requires one more year of data. Thus one advantage of the longitudinal estimator disappears in the presence of measurement error. With four or more repeated cross-sections, the model is obviously overidentified and hence subject to test.

7.6.8. Covariance stationary errors

For simplicity, we implicitly condition on X (see Heckman and Robb, 1985a, 1986a, for the case in which regressors are present.) For any model with stationary errors

[70] Notice that a test that the numerator is zero is a test that $\rho = 1$. Thus one can test the identifying condition that $\rho \neq 1$.

$$\mathrm{Var}(Y_t) = \alpha^2(1-P)P + 2\alpha \mathrm{E}(U_t \mid D=1)P + \sigma_U^2, \qquad \text{for } t > k,$$

$$\mathrm{Var}(Y_{t'}) = \sigma_U^2, \qquad \text{for } t' < k,$$

$$\mathrm{Cov}(Y_t, D) = \alpha\, P(1-P) + \mathrm{E}(U_t \mid D=1)P.$$

Note that $\mathrm{E}(U_t^2) = \mathrm{E}(U_{t'}^2)$ by virtue of our assumption of stationarity. Then[71]

$$\alpha = [P(1-P)]^{-1}[\mathrm{Cov}(Y_t, D) - \{[\mathrm{Cov}(Y_t, D)]^2 - P(1-P)[\mathrm{Var}(Y_t) - \mathrm{Var}(Y_{t'})]\}^{1/2}].$$

Replacing sample moments with population moments defines the estimator. Different features of the covariance stationarity assumptions are being exploited. The longitudinal procedure only requires that $\mathrm{E}(U_t U_{t-j}) = \mathrm{E}(U_{t'} U_{t'-j})$ for $j > 0$; variances need not be equal across periods. The repeated cross-section analog above requires only that variances be stationary; covariances could differ among equispaced pairs of the U_t. With more than two cross-sections, the covariance stationarity assumption is overidentifying and hence subject to test.

7.6.9. The anomalous properties of first difference or fixed effect models

Almost all of the estimators considered in this chapter require a comparison group (i.e., a sample of non-trainees). The only exception is the fixed effect estimator in a time homogeneous environment where $\beta_t = \beta_{t'}$. In this case, if condition (6.8′) holds, and if we let $X_{it}\beta = \beta_t$ to simplify the exposition, then

$$\alpha = \mathrm{E}(Y_t \mid D=1) - \mathrm{E}(Y_{t'} \mid D=1).$$

No information on non-participants is needed, although information on participation or non-participation by the same persons is required.[72] This is not a general feature of the other estimators that we have considered. Even in stationary environments, other estimators require both participants and non-participants. Even the fixed effect estimator requires information on non-participants in a non-stationary environment.

Many of the procedures considered here can be implemented using only post-program data. It is not necessary to have pre-program background data. The covariance stationarity estimators of Section 7.6.4, certain repeated cross-section estimators, and first difference methods constitute exceptions to this rule. In this sense, those estimators are anomalous.

Fixed effect estimators also are robust to departures from the random sampling assumption. For instance, suppose condition (6.8′) is satisfied, but that the available data oversample or undersample trainees (i.e., the proportion of trainees in the sample does not

[71] The negative root of the quadratic equation for α derived from the three moments presented in the text does not identify the parameter. For details, see Heckman and Robb (1985a).

[72] Strictly speaking, we can implement the estimator by sampling participants in the same cohorts without sampling the same persons in t and t'. Recall our discussion of the repeated cross-section estimators.

converge to $P = \mathrm{E}(D)$). Suppose further that the analyst does not know the true value of P. Nevertheless, a first difference regression continues to identify α. As noted in Section 7.7, many other procedures do not share this property.

7.6.10. Robustness of panel data methods in the presence of heterogeneous responses to treatment

It is not surprising that estimators that exploit properties of covariances and variances of model residuals are affected by changes in the properties of the residuals. We have already noted in Section 3 that when responses to treatment are heterogeneous, we acquire a non-standard error term (see (3.7) and (3.9)). As we move from the common coefficient case to the heterogeneous-response case, we encounter new phenomena. Some of the estimators we have considered are robust to the introduction of heterogeneity. Others are not.

In this chapter, we focus on estimating the impact of treatment on the treated so Eq. (3.9) and its error term are the appropriate objects of attention. The induced heteroscedasticity clearly makes the repeated cross-section estimator based on stationarity invalid whether or not $U_{1t} - U_{0t}$ is anticipated in making program participation decisions. Without modification the longitudinal estimator based on covariance stationarity also is invalid.

In contrast, the fixed effect estimator (applied to panels or repeated cross-sections) is robust to heterogeneity in responses provided that the object of an evaluation is to identify $\mathrm{E}(Y_{1t} - Y_{0t} \mid X, D = 1)$. To see this point notice that the fixed effect estimators (for panels or repeated cross-sections) only use conditional mean properties of the errors. From the definition of the parameter (3.8) and Eq. (3.9), the error component induced by heterogeneous responses has mean zero ($\mathrm{E}[(U_1 - U_0) - \mathrm{E}(U_1 - U_0 \mid X, D = 1) \mid X, D = 1] = 0$). Thus the properties of the estimator are not affected by heterogeneity in response when treatment on the treated is the parameter of interest. The selection problem arises solely from dependence between U_0 and D.

The autoregressive estimators provide an interesting example where the introduction of response heterogeneity affects the panel data version the estimator for the effect of treatment on the treated but not the repeated cross-section version. We develop our analysis of the autoregressive estimator in this context in two stages. First assume that the difference in the outcomes in any two periods is time invariant:

$$Y_{1t} - Y_{0t} = \alpha \qquad t > k.$$

Then letting

$$U_{0t} = \rho U_{0,t-1} + \varepsilon_t,$$

where ε_t is independent and identically distributed (i.i.d), we may now write the outcome equation as

$$Y_t = [X_t - X_{t'}\rho^{t-t'}]\beta + (1 - \rho^{t-t'})D\mathrm{E}(\alpha \mid X, D = 1)$$

$$+\rho^{t-t'}Y_{t'} + \sum_{j=0}^{t-(t'+1)} \rho^j \varepsilon_{t-j} + (1-\rho^{t-t'})D[\alpha - \mathrm{E}(\alpha \mid X, D=1)]. \tag{7.29}$$

Observe that even if the ε_t do not determine program participation for periods $t > k$,[73]

$$\mathrm{Cov}(Y_{t'}, D(\alpha - \mathrm{E}(\alpha \mid X, D=1))) \neq 0.$$

Consequently $Y_{t'}$ is correlated with the error term in the model and additional identifying information is required. But recall that the repeated cross-section estimator only uses group means. Just as in the case of the fixed effect estimator, the final error component of Eq. (7.29) averages out to zero when means are constructed. Thus the repeated cross-section estimator is robust to the introduction of response heterogeneity in the model, while the panel data version of the estimator is not.

This point is more general. All repeated cross-section estimators based on means that identify the parameter in the case of a common effect are consistent for $\mathrm{E}(Y_{1t} - Y_{0t} \mid X, D=1)$ in the random coefficient case. The new error component introduced when responses to treatment are heterogeneous averages out over persons. This is a property of the additive separability that underlies the entire class of estimators examined in this section of the chapter and clearly demonstrates the dependence of the properties of these estimators on functional form assumptions.

For the more general autoregressive processes given by

$$U_{1t} = \rho_1 U_{1,t-1} + \varepsilon_{1t},$$

$$U_{0t} = \rho_0 U_{0,t-1} + \varepsilon_{0t},$$

where $\mathrm{E}(\varepsilon_{1t}) = \mathrm{E}(\varepsilon_{0t}) = 0$, and $(\varepsilon_{1t}, \varepsilon_{0t})$ is i.i.d across persons but $\mathrm{E}(\varepsilon_{1t}\varepsilon_{0t}) \neq 0$, the autoregressive estimator is no longer clearly defined. The parameter "treatment on the treated" is now, in general, period dependent, even if $\beta_{1t} = \beta_{0t} = \beta$, because $\mathrm{E}(U_{1t} - U_{0t} \mid X, D=1)$ depends on period t. In addition, unless $\rho_1 = \rho_0$, it is no longer true that we can exploit the trick used to obtain Y_t in terms of lagged $Y_{t'}$ and eliminate $U_{t'}$ in (7.1a) to (7.1c).

When $\rho_1 = \rho_0 = \rho$, U_t still can be written in autoregressive form:

$$U_t = DU_{1t} + (1-D)U_{0t} = \rho^{t-t'}[DU_{1t'} + (1-D)U_{0t'}] + \sum_{j=0}^{t-(t'+1)} \rho^j [D\varepsilon_{1,t-j} + (1-D)\varepsilon_{0,t-j}].$$

Assuming a common β in both regimes except for a time-invariant difference in intercepts

[73] Observe that the error term for $Y_{t'}$ includes $D(\alpha - \mathrm{E}(\alpha \mid X, D=1))$. Then for the variable coefficient model the final term in (7.29) is correlated with this component of $Y_{t'}$. The covariance is $(1-\rho^{t-t'})\mathrm{E}(D(\alpha - \mathrm{E}(\alpha \mid X, D=1))^2) = (1-\rho^{t-t'})\mathrm{Var}(\alpha \mid X, D=1)P \neq 0$ where $P = \Pr(D=1)$. The phenomenon here is similar to the fixed effect bias analyzed by Balestra and Nerlove (1966) except that the fixed effect in our model is state contingent: D times the fixed effect $\alpha - \mathrm{E}(\alpha \mid X, D=1)$.

α, the parameter treatment on the treated for period t is

$$[\alpha + \mathrm{E}(U_{1t} - U_{0t} \mid X, D = 1)] = \alpha + \rho^{t-t'}\mathrm{E}(U_{1t'} - U_{0t'} \mid X, D = 1)$$

$$+ \sum_{j=0}^{t-t'} \rho^j \mathrm{E}(\varepsilon_{1,t-j} - \varepsilon_{0,t-j} \mid X, D = 1). \tag{7.30}$$

Applying the autoregressive transform, we obtain

$$Y_t = (X_t - X_{t'}\rho^{t-t'})\beta + [\alpha(1 - \rho^{t-t'}) + \sum_{j=0}^{t-(t'+1)} \rho^j \mathrm{E}(\varepsilon_{1,t-j} - \varepsilon_{0,t-j} \mid X, D = 1)D + \rho^{t-t'}Y_{t'}$$

$$+ \sum_{j=0}^{t-t'} \varepsilon_{0,t-j} + D\left[\sum_{j=0}^{t-(t'+1)} \rho^j[\varepsilon_{1,t-j} - \varepsilon_{0,t-j} - \mathrm{E}(\varepsilon_{1,t-j} - \varepsilon_{0,t-j} \mid X, D = 1)]\right].$$

If the $\varepsilon_{l,t}$, $t > k$, $l = 0$ or 1, are not forecastable at $t = k$, then the parameter treatment on the treated in period $t > k$ is

$$\alpha + \mathrm{E}(U_{1t} - U_{0t} \mid X, D = 1) = \alpha + \rho^{t-k}\mathrm{E}(U_{1k} - U_{0k} \mid X, D = 1).$$

All of the innovations after $t > k$ are independent of D and hence α and ρ can be identified, as before, by least squares.

If the $\varepsilon_{j,t}$ are evenly partly forecastable by the agents being analyzed, then the final component of the error is correlated with $Y_{t'}$:

$$\mathrm{Cov}\left(Y_{t'}, D\sum_{j=0}^{t-t'} \rho^j(\varepsilon_{1,t-j} - \varepsilon_{0,t-j} - \mathrm{E}(\varepsilon_{1,t-j} - \varepsilon_{0,t-j} \mid X, D = 1))\right) \neq 0.$$

Since two different errors appear in the earnings stream for the $D = 1$ and $D = 0$ choices, they do not difference out as they do in the common coefficient case. In this case, the panel data form of the estimator is inconsistent for the parameter: it is necessary to instrument $Y_{t'}$.

In the general case, with $\rho_1 \neq \rho_0$, the autoregressive estimator breaks down. Different components of the error term decay at different rates, and it is no longer possible to simultaneously eliminate $U_{0t'}$ and $D(U_{1t'} - U_{0t'})$ by one substitution. Thus the method is not in general robust to heterogeneous responses. This lack of robustness to heterogeneous responses is a general feature of many of the panel data estimators discussed in Heckman and Robb (1986a).

7.6.11. Panel data estimators as matching estimators

The simple before–after estimator can be written as a matching estimator using the weighting scheme introduced in Section 7.4.1. To begin, accept assumption (4.A.1) as

valid. For person i at time $t > k$ (k is the program participation period in the notation of Section 4) who has participated in the program, the match is with himself/herself in period $t' < k$. Assume a stationary environment. Letting the match partner be the same individual at time $t' < k$, we match $Y_{0,i,t'}$, $t' < k$ to obtain the following:

$$Y_{1,i,t} - W(i,t')Y_{0,i,t'}, \qquad \text{for } t' < k,$$

where the weight $W(i,t') = 1$. More generally if we have access to more than one preprogram observation per person, one can weight the various terms by functions of the variances determined using the optimal weighting schemes in minimum distance estimation (see Heckman, 1998c, for details.) Thus the comparison group for person i at time t is a weighted average of the available observations for that person over the pre-program observation period:

$$Y^c_{0,i,t} = \sum_{j=0}^{k-1} W(i,j)Y_{0,i,j}, \qquad \text{for } j < k, \tag{7.31}$$

where

$$\sum_{j=0}^{k-1} W(i,j) = 1.$$

Each post-program period can be matched in this way with the pre-program observations. The weights can be chosen to minimize the variance in the sum of the contrasts. (Heckman, 1998c). Assuming that the same treatment effect characterizes all post-program periods, and summing over all post-program observations, we can estimate the treatment on the treated parameter by the sample analog of

$$\sum_{t=k+1}^{T} (Y_{1,i,t} - Y^c_{0,i,t})\varphi(i,t),$$

where

$$\sum_{t=k+1}^{T} \varphi(i,t) = 1$$

and $\varphi(i,t)$ are weights chosen to minimize the variance of this expression. If the treatment effects are different for each post-program period, there is no point in summing across post-program periods.

There is no necessary reason why the weights should be the same on the components. Thus we may write

$$\sum_{t=k+1}^{T} (\alpha(i,t)Y_{1,i,t} - \beta(i,t)Y^c_{i,t}),$$

provided that

$$\sum_{t=k+1}^{T} \alpha(i,t) = 1 \qquad \text{and} \qquad \sum_{t=k+1}^{T} \alpha(i,t) = \sum_{t=k+1}^{T} \beta(i,t),$$

for all i. These conditions enable us to difference out common components and retain identification of $\mathrm{E}(\alpha \mid X, D = 1)$.

If there are trends operating on participants, it is necessary to eliminate them to estimate the parameter of interest. If the trends are common across participants, we are led to using the differences-in-differences method as long as assumption (4.A.2) is valid. In this setting, it is necessary to use a group of persons who do not receive treatment. Accordingly, we can think of creating a comparison person i' for treatment person i:

$$Y_{0,i',t} - \sum_{j=1}^{k-1} W(i',j) Y_{0,i',j}, \qquad \text{for } t > k > j,$$

where

$$\sum_{j=1}^{k-1} W(i',j) = 1 \qquad \text{and} \qquad W(i,j) = W(i',j),$$

for all i,i' and j. This transforms the comparison group to be conformable with the treatment group. We thus create a pairing $i \to i'$, such that persons i and i' have the same weights, i is in the treatment group and i' is in the comparison group, and we can form the difference-in-differences estimator for person i paired with person i' as follows:

$$\left[Y_{1,i,t} - \sum_{j=1}^{k-1} W(i,j) Y_{0,i,j} \right] - \left[Y_{0,i',t} - \sum_{j=0}^{k-1} W(i',j) Y_{0,i',j} \right] \tag{7.32}$$

and $W(i,j) = W(i',j)$ for any (i,i') and all j and where

$$\sum_{i} W(i,j) = 1 \qquad \text{and} \qquad \sum_{i'} W(i',j) = 1.$$

This procedure eliminates common trends and weights the comparison group and treatment group symmetrically. Different weights are required for models with different serial correlation properties (Heckman, 1998c).

More generally, we can form other pairings in the comparison group and compare i to an entire collection of non-treated persons who are operated on by a common trend. For example, we can form an alternative difference-in-differences estimator as follows:

$$\left[Y_{1,i,t} - \sum_{j=0}^{k-1} W(i,j) Y_{0,i,j} \right] - \frac{1}{N_c} \sum_{i'=1}^{N_c} \left[Y_{0,i',t} - \sum_{j=0}^{k-1} W(i',j) Y_{0,i',j} \right] \varphi(i'), \tag{7.33}$$

where N_c is the number of persons in the comparison sample, $\varphi(i')$ is a weight and where

$$\frac{1}{N_c}\sum_{i'=1}^{N_c}\varphi(i') = 1 \qquad \text{and} \qquad \frac{1}{N_c}\sum_{i'=1}^{N_c} W(i',j)\varphi(i') = W(i,j).$$

Difference (7.33) eliminates age- or period-specific common trends or year effects. We can form variance weighted versions of (7.33) to pool information across i to estimate $\mathrm{E}(Y_1 - Y_0 \mid X, D = 1)$ efficiently if the effect is constant (see Heckman, 1998c).

The same scheme can be used to estimate models with person-specific, time-varying variables. Time-invariant variables are eliminated by subtraction. Consider the before–after estimator. Let $A_{it}(Y_{it})$ be an "adjustment" to Y_{it}, where

$$A_{it}(Y_{i,t}) = Y_{i,t} - g(X_{i,t}).$$

Then the comparison group for person i based on his preprogram adjusted outcomes can be written as

$$A_{it}^{c}(Y_{i,t}) = \sum_{j=0}^{k-1} W(i,j)A_{jt}(Y_{0,i,j})$$

and the before–after estimator can now be written in terms of adjusted outcomes as follows:

$$A_{it}(Y_{1,i,t}) - A_{i,t}^{c}(Y_{i,t}).$$

We can make a similar modification to the difference-in-differences scheme:

$$\left[A_{it}(Y_{1,i,t}) - \sum_{j=0}^{k-1} W(i,j)A_{jt}(Y_{0,j,t})\right] - \left[A_{i',t}(Y_{1,i',t}) - \sum_{j=0}^{k-1} W(i',j)A_{i',j}(Y_{0,i',t})\right],$$

where $W(i,j) = W(i',j)$ for all i,i', and

$$\sum_{j=0}^{k-1} W(i',j) = 1 \qquad \text{and} \qquad \sum_{j=0}^{k-1} W(i,j) = 1.$$

This modification eliminates non-invariant components. This enables us to generalize the simple before-after estimator to a case where person-specific and period-specific shocks operate on agents. This produces a large class of longitudinal estimators as special cases of the weighting scheme introduced in our discussion and is the basis for a unified treatment of a variety of evaluation estimators. Heckman (1998a) presents a comprehensive analysis and many examples of weights for different traditional econometric estimators.

7.7. Robustness to biased sampling plans

Virtually all estimation methods can be readily adjusted to account for choice-based sampling (i.e., oversampling of trainees relative to comparison group members) or

measurement error in training status among the comparison group (some of the comparison group members have taken training). Some methods require no modification at all.

The data available for analyzing the impact of training on earnings are often nonrandom samples. Frequently they consist of pooled data from two sources: (a) a sample of trainees selected from program records and (b) a sample of non-trainees selected from some national sample. Typically, such samples overrepresent trainees relative to their proportion in the population. This creates the problem of choice-based sampling first analyzed in a more general form by Rao (1965, 1986) and applied by Manski and Lerman (1977) and Manski and McFadden (1981).

A second problem, contamination bias, arises when the training status of certain individuals is recorded with error. Many control samples such as the US Current Population Survey or the US Social Security Work History data do not reveal whether or not persons have received training. These sampling situations combine the following types of data:

(A) outcomes, observable characteristics and participation status for a sample of trainees ($D = 1$);

(B) outcomes, observable characteristics and participation status for a sample of non-trainees ($D = 0$);

(C) outcomes and observable characteristics for a national comparison sample of the population (e.g., CPS or Social Security records) where the training status of persons is not known.If type (A) and (B) data are combined and the sample proportion of trainees does not converge to the population proportion of trainees, the combined sample is a choice-based sample. If type (A) and (C) data are combined with or without type (B) data, there is contamination bias because the training status of some persons is not known.

We can modify most procedures developed in the context of random sampling to consistently estimate $\mathrm{E}(\alpha \mid X, D = 1)$ using choice-based samples or contaminated comparison groups. In some cases, a consistent estimator of the population proportion of trainees is required. We illustrate these claims by showing how to modify the instrumental variables estimator to address both sampling schemes. We briefly consider several other methods as well. Heckman and Robb (1985a, 1986b) give explicit case-by-case treatment of these issues for a variety of estimators including all of the panel data estimators considered in this paper.

7.7.1. The IV estimator and choice-based sampling

If condition (7.17b) is strengthened to read

$$\mathrm{E}(U_0 \mid X, Z, D) = \mathrm{E}(U_0 \mid X), \qquad \text{for } D = 0, 1, \tag{7.17b$'$}$$

the IV estimator is consistent for $\mathrm{E}(\alpha \mid X, D = 1)$ in choice-based samples. The important point to notice is that identification condition (7.17b) is written for the population. By contrast, (7.17b$'$) is written for a subset of the population conditional on $D = 1$ or $D = 0$. If we reformulate the IV condition to apply to the $D = 0$ and $D = 1$ subpopulations, it does not matter how we reweight the subpopulations to form samples – the orthogonality conditions apply to any combinations of them.

To see how to form consistent estimators under the assumptions of Section 7.4.3, let D^* be the event that "a trainee is observed in a choice-based sample." In a sample generated by choice-based sampling, the probability of participation $\Pr(D^* = 1) = P^* \neq P = \Pr(D = 1)$, where P is the probability of participation in the case of random sampling.

Now in the choice-based sample, let U_0^* be the random variable U_0 generated from choice-based sampling, so that

$$\mathrm{E}(U_0^* \mid X, Z) = \mathrm{E}(U_0 \mid X, Z, D^* = 1)P^* + \mathrm{E}(U_0 \mid X, Z, D^* = 0)(1 - P^*).$$

If (7.17b′) applies, then we can write

$$\mathrm{E}(U_0^* \mid X, Z) = \mathrm{E}(U_0 \mid X, Z)P^* + \mathrm{E}(U_0 \mid X, Z)(1 - P^*) = \mathrm{E}(U_0 \mid X, Z).$$

Provided P is known, it is possible to reweight the data to secure consistent IV estimators for $\mathrm{E}(\alpha \mid X, D = 1)$ under the assumptions of Section 7.4.3. Simply multiply both dependent and independent observations by the weight

$$\omega = D\frac{P}{P^*} + (1 - D)\left(\frac{1 - P}{1 - P^*}\right)$$

and apply IV to the transformed data. This weighting ensures that (7.17b) applies to the reweighted data. The IV method applied to the reweighted samples consistently estimates the parameters of interest provided that other identifying assumptions are maintained (see Heckman and Robb, 1985a, 1986a).

7.7.2. The IV estimator and contamination bias

For data of type (C), D is not observed. Applying the IV estimator to pooled samples of type (A) and (C) data assuming that all observations in the type (C) data have $D = 0$ produces an inconsistent estimator if the type (C) data includes some trainees. However, with a minimal amount of additional information, it is possible to identify the estimator in this case.

In terms of the IV Eqs. (7.18) or (7.20), it is possible to generate $\mathrm{E}(Y \mid X, Z)$ from the type (C) sample. The type (A) data yield the sample joint distribution of (Y,X,Z) given $D = 1$ and in particular the joint distribution $f(X, Z \mid D = 1)$. Since we know

$$f(X, Z) = f(X, Z \mid D = 1)P + f(X, Z \mid D = 0)(1 - P),$$

we can solve for $f(X, Z \mid D = 0)$ if we know P. From Bayes' rule, we can write (denoting "f" as the density)

$$\Pr(D = 1 \mid X, Z) = \frac{f(X, Z, D = 1)}{f(X, Z)}.$$

The two densities can be constructed from the information in the type (C) and type (A) samples. Thus with knowledge of P, it is possible to estimate $\Pr(D = 1 \mid X, Z)$ for each person and hence to construct the IV estimator for contaminated samples. One can think of this procedure as a data imputation exercise. See Heckman and Robb (1985a,

1986a), Imbens and Lancaster (1996) and Heckman (1998a) for the econometric details.

7.7.3. Repeated cross-section methods with unknown training status and choice-based sampling

The repeated cross-section estimators discussed in Section 7.6.5 are inconsistent when applied to choice-based samples unless additional conditions are assumed.[74] For most of the repeated cross-section estimators, it is necessary to know the identity of the trainees to weight the sample back to the proportion of trainees that would be produced by a random sample to obtain consistent estimators. Hence, the class of estimators that does not require knowledge of individual training status is not robust to choice-based sampling.

Some of the estimators that we have examined are robust to choice-based sampling. Any estimator that is constructed conditional on D has the property of being robust to choice-based sampling. (Recall our discussion of instrumental variables estimators where the condition (7.17b) was modified to hold conditionally on D.) A control function estimator constructs

$$\mathrm{E}(U_{1t} \mid X, Z, D), \tag{7.34a}$$

$$\mathrm{E}(U_{0t} \mid X, Z, D), \tag{7.34b}$$

and works with the purged residuals

$$U_{1t} - \mathrm{E}(U_{1t} \mid X, Z, D)$$

and

$$U_{0t} - \mathrm{E}(U_{0t} \mid X, Z, D)$$

from the original model. Then the parameters of (7.34a) and (7.34b) are estimated along with the remaining parameters of the model. Identification conditions for control function models are given in Heckman and Robb (1985a).[75] The selection bias terms $K_0(P(Z))$ and $K_1(P(Z))$ in Eqs. (7.16a) and (7.16b) are examples of control functions with the inverse Mills' ratio as the leading example used in empirical work. Likewise, the autoregressive estimator of Heckman and Wolpin (1976) discussed in Section 7.6.3 is a control function estimator where

$$K_t = \rho^{t-t'} U_{t'}, \qquad \text{for } t > t' > k$$

and where $Y_{t'} - \beta_{t'} - \alpha D = U_{t'}$. The higher-order autoregression schemes discussed in

[74] This is not always true. For example, when the environment is time homogeneous, $(\bar{Y}_t - \bar{Y}_{t'})/P$ remains a consistent estimator of $\mathrm{E}(\alpha \mid X, D = 1)$ in choice-based samples as long as the same proportion of trainees are sampled in periods t' and t.

[75] They present conditions under which it is possible to identify the control functions apart from the parameters of the model. See also Heckman (1990).

Heckman and Robb (1985a, p. 223) are also control functions. They discuss additional control functions based on factor models and optimal forecasting schemes.

The basic principle of the control function is that of constructing conditional means of the errors in each regime ($D = 0, 1$) and estimating these conditional means and the other parameters of the model. As long as the control function is defined to be conditional on D, the control estimator is robust to choice-based sampling.

7.8. Bounding and sensitivity analysis

Since the problem of "causal analysis" is intrinsically a missing data problem, methods from the missing data literature can be used to solve the problem of causal inference, and to provide bounds on the missing data. Various bounding schemes proposed in the recent literature can be regarded as applications of the 1970s and 1980s literature on missing data.

The prototype for this approach is presented in a paper by Smith and Welch (1986) who consider both a sensitivity analysis and a bounding analysis in examining the effect of selection bias on the measured wage of blacks. Commenting on a paper by Butler and Heckman (1977), who attribute some part of the growth in black real wages observed in the US in the 1960s to selective withdrawal of the least skilled workers from the labor force, Smith and Welch (1986) apply the law of iterated expectations to write the true wage of all blacks $\mathrm{E}(W_B)$ as

$$\mathrm{E}(W_B) = \mathrm{E}(W_B \mid L_B = 1)P(L_B = 1) + \mathrm{E}(W_B \mid L_B = 0)P(L_B = 0), \tag{7.35}$$

where $\mathrm{E}(W_B \mid L_B = 1)$ is the wage of black workers, $\mathrm{E}(W_B \mid L_B = 0)$ is the wage of black labor force dropouts would have received if they would have worked and $\mathrm{E}(W_B)$ is the mean wage of all blacks.[76] $P(L_B = 1)$ is the proportion of the black population that is working. Observed (consistently estimable) are $\mathrm{E}(W_B \mid L_B = 1)$ and $P(L_B = 1)$ (and hence $1 - P(L_B = 1)$). Missing data on the wages of non-participants make $\mathrm{E}(W_B \mid L_B = 0)$ non-identified and hence $\mathrm{E}(W_B)$ is non-identified.

Smith and Welch (1986) adopt several solutions to this identification problem which have been widely applied in the evaluation literature. The first is to use panel data to follow non-workers over time and find the wage that is observed most recently to replace the missing wage. The second is to bound the missing parameter $\mathrm{E}(W_B \mid L_B = 0)$ assuming that $[\mathrm{E}(W_B \mid L_B = 0) = \gamma \mathrm{E}(W_B \mid L_B = 1)]$ for $0.5 \leq \gamma \leq 1$. By varying γ over a range of values, they perform a sensitivity analysis or bounding analysis that has recently become fashionable in applied social science. Their methods apply directly to the selection problem. Suppose we know $\mathrm{E}(Y_0 \mid D = 0)$. We seek to know $\mathrm{E}(Y_0 \mid D = 1)$ to construct the counterfactual $\mathrm{E}(Y_1 - Y_0 \mid D = 1)$. By using bounds connecting $\mathrm{E}(Y_0 \mid D = 0)$ to $\mathrm{E}(Y_0 \mid D = 1)$, it is possible to bound $\mathrm{E}(Y_1 - Y_0 \mid D = 1)$. (Recall that $\mathrm{E}(Y_1 \mid D = 1)$ is known).

Glynn and Rubin (1986) present a similar analysis of what they call "mixture models."

[76] We use a simplified notation to convey the main idea in their work.

Like Smith and Welch (1986), they analyze two cases: (a) one where the missing data in one period can be obtained in another period and (b) one where they perform a "sensitivity" analysis by varying the unidentified parameters of the model. Rosenbaum (1995) summarizes a series of papers going back to the late 1950s that bound estimated causal effects by bounding the range of the unobserved parameters of the model.

In this section of the chapter, we draw on the comprehensive analysis of Balke and Pearl (1993, 1997), Balke (1995), and Chickering and Pearl (1996) on bounding causal parameters. Using linear programming methods, they extend the work of Robins (1989) and Manski (1995) to present the tightest possible *non-parametric* bounds for causal parameters. These methods exploit certain classical inequalities of probability theory. Instead of analyzing a model with a high level of generality, consider a specific model of missing data that links recent analyses of bounds for causal parameters to the classical problem of missing data in contingency analysis. The Holland (1986, 1988) and Rubin (1974, 1978) model is essentially one for a contingency table with missing data. Results in the literature on missing data in the contingency tables apply directly to the model of causal effects.

Fig. 7 considers a model of potential outcomes for each person i when there are two possible values for each potential outcome $Y_0 \in \{0,1\}$, $Y_1 \in \{0,1\}$, $D \in \{0,1\}$. This produces a $2 \times 2 \times 2$ table. In the case of a randomized experiment where randomization is done after persons have attempted to enroll in the program, the row and column margins of the left ($D = 1$) table are known but not the individual cells. One piece of identifying information is missing. A monotonicity assumption (e.g., $P_{101} = 0$) fully identifies the table. This assumption says that among the persons who enter the program, there are no persons who would switch from 1 to 0 status. One can use the Frechet bounds to obtain ranges of possible values, using the column and row marginal distributions for the table (see Heckman and Smith, 1993; Heckman et al., 1997c, for discussions and the first applications of these bounds to the evaluation problem).[77] These bounds produce the tightest possible bounds on the elements of a contingency table given the marginal distributions. In practice, these bounds are usually very wide as those authors, and the vast literature in statistics that precedes them, have shown.

The more general case with observational data is one where the column totals are known for the $D = 1$ table and the row totals are known for the $D = 0$ table. The remaining elements are not known.

[77] For any joint distribution for discrete or continuous random variables, $F(a,b)$, with marginal distributions $F(a)$ and $F(b)$, $\text{Max}[F(a) + F(b) - 1, 0] \leq F(a,b) \leq \text{Min}[F(a), F(b)]$. The upper bound is a trivial consequence of the fact that $\Pr(A \leq a \cap B \leq b) \leq \text{Min}(\Pr(A \leq a); \Pr(B \leq b))$. The lower bound is equally straightforward to derive. Partition the space (A,B) into four mutually exclusive regions: $R_1 = (A \leq a, B \leq b)$, $R_2 = (A \leq a, B > b)$, $R_3 = (A > a, B \leq b)$, $R_4 = (A > a, B > b)$, where (*) is defined as $\Pr(R_1) + \Pr(R_2) + \Pr(R_3) + \Pr(R_4) = 1$. Observe that $R_1 \cup R_2 = (A \leq a)$ while $R_1 \cup R_3 = (B \leq b)$. $\Pr((R_1 \cup R_2) \cup (R_1 \cup R_3)) = \Pr(R_1 \cup R_2 \cup R_3) = 1 - \Pr(R_4)$. (**) $\Pr(R_1 \cup R_2) + \Pr(R_1 \cup R_3) = \Pr(A \leq a) + \Pr(B \leq b)$. Subtracting (**) from (*) and rearranging, we obtain $\Pr(R_1) = \Pr(A \leq a) + \Pr(B \leq b) - 1 + \Pr(R_4)$. Since $\Pr(R_4) \geq 0$, $\Pr(R_1) \geq \Pr(A \leq a) + \Pr(B \leq b) - 1$ so $F(a,b) \geq F(a) + F(b) - 1$ but since probabilities cannot go negative $F(a,b) \geq \max(0, F(a) + F(b) - 1)$.

Consider bounds for the treatment on the treated parameter TT: $\mathrm{E}(Y_1 - Y_0 \mid D = 1)$. In terms of cell proportions:

$$TT = \mathrm{E}(Y_1 - Y_0 \mid D = 1) = \frac{P_{.11} - P_{1.1}}{P_{..1}} = \frac{P_{011} - P_{101}}{P_{..1}}.$$

For the case of observational data the solution is straightforward. The linear program to bound the parameter is

Max TT subject to

$$\hat{P}_{.01} = P_{001} + P_{101} \qquad \text{(columns determined)} \tag{7.36a}$$

$$\hat{P}_{.11} = P_{011} + P_{111} \tag{7.36b}$$

and

$$\hat{P}_{0.0} = P_{000} + P_{010} \qquad \text{(rows determined)} \tag{7.36c}$$

$$\hat{P}_{1.0} = P_{100} + P_{110}. \tag{7.36d}$$

We are free to make P_{011} maximal by setting $P_{111} = 0$ (so $\hat{P}_{.11} = P_{011}$) and to make P_{101} minimal by setting $P_{001} = \hat{P}_{.01}$. No constraints are violated because we have freedom to pick the row totals in the $D = 1$ table. By the same token we can make P_{011} minimal by setting $P_{111} = \hat{P}_{.11}$ so $P_{011} = 0$ and make P_{101} maximal by setting $P_{001} = 0$ and $\hat{P}_{0.1} = P_{1.1}$.

$$\Pr(Y_1 = 1 \mid D = 1) = \frac{\hat{P}_{.11}}{P_{..1}} \geq TT \geq -\frac{\hat{P}_{.0}}{P_{..1}} = -\Pr(Y_1 = 0 \mid D = 1).$$

Access to experimental data sharpens these bounds to a point. In this case, we know both the row totals and the column totals of the $D = 1$ table. We supplement linear inequalities (7.36a) and (7.36b) by

$$\hat{P}_{0.1} = P_{001} + P_{011}, \tag{7.36e}$$

$$\hat{P}_{1.1} = P_{101} + P_{111}. \tag{7.36f}$$

Now the formal optimization problem is apparently harder; Max TT subject to (7.36a), (7.36b) and (7.36e), (7.36f). Using (7.36a) and (7.36b) we obtain

$$\hat{P}_{.11} - \hat{P}_{1.0} = P_{011} - P_{101}$$

so the parameter is exactly identified. Using the Balke–Pearl methods, we can bound any parameter, or any empty cell in a contingency table analysis, using linear programming methods.

It is important to recognize that these are *non-parametric* bounds. They *do not* capture

the full potential variability in the estimated parameter values when parametric structure is imposed on the P, as is commonly done in applied work. Nor do they capture uncertainty about the X. To get the full range of variability in the parameter requires solving the non-linear program across models M and possible regressors $\bar{\bar{X}}$ used to generate the $P_{ijk}(x,m)$, where x is a choice of regressors and m is the particular model. A full characterization of model variability in this framework is given by choosing that m and x that maximize

$$\frac{P_{011}(x,m)}{P_{\cdot\cdot 1}(x,m)} - \frac{P_{101}(x,m)}{P_{\cdot\cdot 1}(x,m)},$$

that is

$$\underset{m\in M, x\in\bar{\bar{X}}}{\text{Max}} \left[\frac{P_{011}(x,m)}{P_{\cdot\cdot 1}(x,m)} - \frac{P_{101}(x,m)}{P_{\cdot\cdot 1}(x,m)} \right]$$

subject to appropriate (i.e., modified for X and m) constraints. These bounds account for model uncertainty and regressor misspecification. A full characterization of this problem remains to be developed.

8. Econometric practice

One of the most important lessons from the literature on evaluating social programs is that choices made by evaluators regarding their data sources, the composition of their comparison groups, and the specification of their econometric models have important impacts on the estimated effects of training. As noted in Section 7, the choice of a comparison sample can affect the statistical properties of an estimator applied to that sample. Under the conditions specified there, for certain comparison groups, simple mean comparisons between treatments and controls identify the parameters of interest.

The purpose of this section is to draw from the empirical literature to show why and how these choices matter. To begin our discussion, we first discuss the types of data used in most evaluations of active labor market policies and show how the source of data affects the impact estimates. Next, we draw on the work of Heckman et al. (1996b, 1998b), who collect unusually rich data compared to what is usually available to program analysts, to analyze the sources of measured selection bias. Their findings provide an informative guide to the construction of datasets for future evaluations.

In the third section, we present a small scale simulation study of alternative evaluation estimators which make different assumptions about program participation decision rules, outcome equations and their interrelationship. This simulation study summarizes the lessons of Section 7 and reveals that no universally valid estimator exists or is ever likely to be found. In the fourth and concluding section, we consider the logic that underlies the use of widely-applied "specification tests" to check the validity of an evaluation model by determining if it "aligns" the earnings (or other measures) of participants and non-participants prior to their enrollment in the program. The method is not guaranteed to pick a

correct evaluation model. We demonstrate the practical importance of this point and show how two different alignments used in the literature produced two very different and controversial impact estimates for the same program.

8.1. Data sources

To evaluate active labor market policies requires choosing data sources from which to construct comparison groups and treatment groups. In this subsection, we discuss these issues and describe the advantages and disadvantages of the various types of data typically used to evaluate employment and training programs. The decision about what data source or data sources to use has important implications for several aspects of an evaluation. In both experimental and non-experimental evaluations, the decision affects how much the evaluation will cost, how large the analysis sample will be (which affects the size of the training effect that can be statistically distinguished), what outcome variables can be studied, the time period over which the outcome variable can be measured and the amount and type of measurement error in the outcome variable. In non-experimental evaluations, the decision also affects which of the non-experimental evaluation methods discussed in this chapter can be used and whether or not the comparison group can be located in the same local labor markets as the participants. By affecting these aspects of an evaluation, the choice of a data source affects the final impact estimates.

A comparison between the studies of Fraker and Maynard (1987) and LaLonde (1986) illustrates that the choice of a data source can vitally affect the impact estimates obtained in a social experiment. Both of these studies examined the National Supported Work Demonstration. The demonstration included one baseline and up to four followup surveys of the treatments and controls. LaLonde (1986) used this survey data for his analysis, while Fraker and Maynard (1987) used administrative data on annual earnings from the US Social Security Administration (SSA). There exist striking differences between the experimental impact estimates reported in the two studies. Using the survey data, the annual impact of Supported Work on the earnings of AFDC (welfare) women was $1641 in 1978 and $851 in 1979. By contrast, when using the SSA earnings data on the same participants and controls, the annual impact was $505 in 1978 and $351 in 1979. The different data sources produce a difference in the estimated experimental impacts of $1135 in 1978 and of $500 in 1979. The sensitivity of the impact estimates to the data used in the analysis is similar in magnitude to their sensitivity to different econometric modelling assumptions and is large enough to affect the conclusions of a cost-benefit analysis.[78]

[78] Similar sensitivity to the choice of data source was found in the National JTPA Study. For male youth estimates using survey data showed a negative and statistically significant impact from the program, while estimates using administrative data from state Unemployment Insurance (UI) records showed essentially a zero impact. See Bloom et al. (1993). Some of the difference between the estimates shown in Table 4 based on the official 18 and 30 month NJS impact reports results from the fact that the 18 month estimates rely only on survey data while the 30 month estimates rely on a combination of survey data and earnings data from state UI records.

8.1.1. Using existing general survey data sets

In non-experimental evaluations, existing survey datasets constitute one potential source from which comparison groups can be drawn. Examples of such datasets in the US include the Current Population Survey, a large cross-sectional survey which was the source of comparison groups for some of the CETA evaluations, or the National Longitudinal Survey of Youth (NLSY), a widely used panel dataset. Such datasets are not generally used to collect information on participants because they usually collect little, if any, information on receipt of public sector training.

The key advantages of using existing datasets as a source for non-experimental comparison groups are cost and sample size. Using an existing dataset avoids the costs of designing, testing and fielding a survey as well as the costs of locating potential comparison group members. General purpose datasets typically have large samples and are available for a modest fee. Depending on the dataset, they may provide either repeated cross-sectional samples, as with the US CPS, or a long panel, as with the NLSY. In general, a large list of regressors is available for subgroup analysis.

Existing survey data have four key disadvantages for evaluation research. First, it is often difficult to construct comparison groups of persons in the same local labor markets as participants from existing datasets due to sample size limitations and constraints imposed by privacy concerns on the level of detailed locational information made available to researchers. As we show in the next section, this is a severe limitation because variation across local labor markets plays a large role in explaining the earnings and employment variation of unskilled workers who are the targets for active labor market policies (Heckman et al., 1998b). Second, in contrast to what is possible when fresh survey data are collected, it is impossible to obtain specific variables of interest for the program being evaluated not already present in the existing data. Such variables might include the detailed information on recent labor force status histories noted as important determinants of program participation in Section 6. Third, because receipt of public training is often not measured or is not measured well in these data, contamination bias becomes an issue (Heckman and Robb, 1985a) as some members of the comparison group are likely to have received the treatment being evaluated. (Recall our discussion in Section 7.7). Finally, using existing datasets to construct a comparison group often entails using different survey instruments with different definitions of the same outcome variable for participants and comparison group members in an evaluation. Comparing outcomes measured in two different ways adds an important potential source of bias to the impact estimates reported for a program (Smith, 1997b).

8.1.2. Using administrative data

Many evaluations of active labor market policies in the US and Scandinavia rely on administrative data. These pre-existing data generally consist of administrative earnings records collected for tax purposes and administrative records on social assistance receipts. They are often combined with administrative data on the receipt of training from program records. The key advantages of such data are the low cost of acquiring them and lack of

certain types of measurement error. The costs are low on several dimensions. The fixed costs of extracting administrative earnings records are typically modest compared to the costs of collecting comparable data from surveys.[79] Moreover, the marginal costs of increasing the sample size or the number of time periods of data obtained are often very small. For example, recent estimates of the marginal cost of obtaining 10 years of quarterly data on an individual's earnings and social assistance receipts are approximately $2.50. These low costs make such data particularly attractive for non-experimental evaluations in which longitudinal methods will be used. Because these earnings data are used for tax purposes, there are strong incentives for authorities to minimize reporting errors for earnings, so they are likely to be much more accurate, for the types of earnings they intend to measure, than earnings data obtained from surveys.

Administrative data also have important limitations in the context of evaluating employment and training programs. First, these data typically consist of quarterly or annual earnings and little else is reported. Monthly earnings, as well as other outcomes of interest such as wage rates, hours worked and employment spells, are nearly always unavailable. Consequently, in the US, where researchers have relied on such data, relatively little research has looked at the impact of training on wages. This outcome is of great theoretical interest, because higher wages for trainees indicate that training raised their productivity. An exclusive focus on earnings or employment rates does not determine what part of the training impact results from increased productivity of the workers as measured by their hourly wage rates and what part results from the displacement of non-trainees in the labor market (Johnson, 1979).

Second, because governments maintain administrative records for tax and benefits purposes, these earnings measures may not equal total earnings. For example, many recent US evaluations use earnings from state unemployment insurance (UI) records. These data include earnings from jobs "covered" by the UI system, but omit earnings from self-employment, from employers in other states, and from sources not covered by the UI system. As a result, administrative earnings measures tend to be lower than those reported by individuals in surveys (Kornfeld and Bloom, 1996; Smith, 1997b).

Finally, administrative data typically contain only very basic information on demographic characteristics. For example, Table 5 shows that Ashenfelter's (1978) study of MDTA, which uses detailed information on annual pre-program and post-program earnings histories from SSA records, includes only very limited demographic information - just age, sex and race. No information on labor force histories, educational levels, training history, family status or geographical location was available in the data. Lack of data on individual characteristics limits the subgroup analyses that evaluators can perform and makes it difficult to justify the application of non-experimental methods such as matching whose plausibility depends on access to a rich set of conditioning variables.

[79] There are exceptions to this rule. In the NJS, state personnel were unable to provide useable unemployment insurance earnings data in 4 of the 16 states containing training centers in the NJS despite repeated attempts.

8.1.3. Collecting new survey data

An alternative to using existing data sources is to collect fresh survey data on participants and on controls or comparison group members. This choice has both advantages and disadvantages. The first advantage relative to using either existing survey or administrative datasets is that the evaluator has complete control over the information collected on the survey, and so can design the survey in light of the variables of interest in the study and, in non-experimental evaluations, in light of the econometric methods to be used. The second advantage relative to using existing data is that the sampling plan for the survey can target comparison group members in the same local labor markets as participants. A third advantage of collecting new survey data is that relative to administrative data, the analyst can obtain additional outcome measures such as wage rates and employment transitions, and can conduct a wider variety of subgroup analyses.

The most important disadvantage of collecting fresh survey data relative to using either administrative data or existing survey datasets is the high cost of doing so. The total costs of collecting new survey data can vary widely depending on whether evaluators obtain these data through a survey sent through the mail, conducted over the telephone, or during a person-to-person interview. Surveys done through the mail are inexpensive, but typically are plagued by very low response rates; surveys conducted in person are expensive but have very high response rates. In some studies more than one type of survey is used to obtain the data. The fixed costs associated with surveys also can vary widely depending on whether evaluators use an existing survey instrument, whether the survey is automated, and whether or not the interviewers require training.

Most program evaluations based on new survey data use either a telephone or in-person survey. Phone surveys are attractive because the marginal cost of obtaining an additional response is relatively low. Such costs, which include the interview, editing, and coding of the data, are approximately $50 per observation. Longer interviews increase these costs modestly. Average costs are generally double this amount or more. Telephone surveys can be problematic especially when surveying low income populations, because response rates are often significantly lower than for in-person interviews. One practical problem in low income populations is that some respondents may not have a working telephone at the time of the survey. If the survey is done in person, the marginal cost of obtaining an additional observation more than doubles. Further, these marginal costs rise sharply with the response rate. Additional respondents become harder to find. Average costs for samples of modest size obtained from in-person surveys range as high as $500 per observation. If evaluators wish to return and resurvey the same sample at a later date, these costs may not fall appreciably in the second wave. Low income populations are often highly mobile and resources must be expended locating persons who have moved.

The costs of collecting new survey data are likely to be lower for program participants than for members of the comparison group. To obtain a sample of participants, evaluators can use the administrative records that contain information such as the individual's address, phone numbers, and sometimes the most recent employer. Such information is advantageous, because locating respondents is an important component of the cost of

surveys. In contrast, obtaining information on a comparison group requires evaluators to construct a list of comparable persons. One criterion for sample inclusion might be to include persons eligible for the program, who did not participate. An advantage of using non-participating program applicants is that they constitute a ready list from which evaluators can sample for their survey. Another method for selecting a comparison group is to first conduct a short "screening survey" to obtain a list of individuals who were eligible for the program, but did not participate. Even in low income neighborhoods, the fraction of respondents found to be eligible is typically low, so evaluators must conduct many short interviews to obtain a sufficient number of comparisons. Even when using a telephone survey, these procedures can double the marginal cost of obtaining an observation for a comparison group member.

Collecting new survey data can be expensive. As a result, there is no reason to expect that careful non-experimental evaluations that collect new survey data are appreciably less costly than experimental evaluations. The marginal cost per participant of administering an experiment is small. The cost of obtaining a high quality comparison group in a non-experimental evaluation can be very high. A dramatic example of the high cost of collecting new survey data in a non-experimental evaluation is the cost of obtaining the non-experimental comparison group used in the NJS. This sample cost $3.5 million (1990) to collect responses from 3000 persons, in two waves, from just four of the 16 sites included in the study (Smith, 1994). Most of these responses were obtained using a telephone survey. The average cost was a little more than 1000 per observation. Particularly large were the costs associated with locating eligible persons not participating in JTPA.[80]

Related to the general issue of cost is an important tradeoff that affects evaluators who collect their own survey data. Researchers often seek longterm followup data on outcomes, to determine whether shortterm program impacts persist. Non-experimental researchers planning to use many of the longitudinal methods considered in Section 7.6 require information on outcomes in periods prior to the decision to participate in training. The marginal cost of obtaining additional periods of outcome data either before or after participation is usually low for administrative data. With survey data, the evaluator must choose between constructing a panel by fielding costly additional surveys, or tolerating the degradation of data quality as the length of the survey recall period increases.[81]

8.1.4. Combining data sources

One solution to the limitations of any particular type of data is to construct a new dataset by combining more than one type of data. Evaluators often combine administrative data on outcomes with survey data on the characteristics of participants and of comparison or control group members. Analysts then have access to relatively rich data on individual

[80] We are grateful to personnel at Mathematica Policy Research, MDRC, NORC, Westat, and the W.E. Upjohn Institute for Employment Research for providing us with information on the cost of collecting survey data.

[81] Bound et al. (1994) provide evidence of recall effects in labor market survey data and Sudman and Bradburn (1982) discuss the general issue of recall in surveys.

regressors as well as a long panel of earnings data that allows implementation of longitudinal estimators of program impact. For example, many of the US CETA evaluations use a dataset that combines program records on trainees with comparison group data drawn from the CPS. The dataset includes matched administrative earnings data from the Social Security Administration for both groups. However, because the comparison group is drawn from the existing CPS dataset, it is not possible to match them to participants in the same local labor markets.

The NJS provides an example of a study which combined new survey data with administrative data. In this evaluation, treatment and control group members completed a baseline survey and one or two followup surveys. These data were combined with administrative earnings data from state UI systems, administrative income data from the US Internal Revenue Service and administrative data on social assistance from state welfare agencies. The NJS also collected both survey and administrative data on its non-experimental comparison group sample. Because the NJS researchers collected fresh survey data rather than using an existing dataset, they were able to locate comparison group members in the same local labor markets as participants.

8.2. *Characterizing selection bias*

We next draw on the work of Heckman et al. (1996b, 1997a, 1998b), and demonstrate the value of better data in conducting evaluations of active labor market policies. Placing people in the same local labor market and administering them the same survey instrument makes an enormous difference to the quality of an evaluation. So does comparing comparable people. We also summarize the best available evidence on the validity of the widely used practice of using "no shows" as a comparison group.

The mean selection bias in using non-participants to approximate participant outcomes conditional on X is given by

$$B(X) = \mathrm{E}(Y_0 \mid X, D = 1) - \mathrm{E}(Y_0 \mid X, D = 0). \tag{8.1}$$

Selective differences in uncontrolled variables (variables on which the analyst cannot condition) produce selection bias. Such differences may arise from self-selection decisions by the agents being studied or from uncontrolled differences between treatments and controls due to the inadequacy of the available data. We argue that much of the bias reported by LaLonde (1986) in his influential study of the effectiveness of econometric estimators arises from the second source – the inadequacy of the data. In ordinary non-experimental evaluations, B is unknown. This produces the evaluation problem. Using data from a social experiment conducted under the conditions specified in Section 5, it is possible to estimate the first term on the right hand side of (8.1). Using a non-experimental comparison group it is possible to estimate the second term.

The conventional measure of selection bias, B, used by LaLonde (1986), Ashenfelter (1978) and Heckman and Hotz (1989) is the mean difference between the earnings of controls and the earnings of comparison group members:

$$B = \mathrm{E}(Y_0 \mid D = 1) - \mathrm{E}(Y_0 \mid D = 0).$$

This is the coefficient on D of a regression of Y_0 on D in a pooled comparison group and control group sample, $Y_0 = \alpha_0 + BD + \tau$ when $\mathrm{E}(\tau \mid D) = 0$. It does not condition on X.

Heckman et al. (1996b, 1998b) estimate the bias term $B(X)$ using non-parametric methods. With their estimated bias, they test the identifying assumptions that justify matching, the classical econometric selection bias estimator and a non-parametric version of difference-in-differences. They show that it is possible to decompose the conventional measure of bias, B, which does not condition on X, into three components. The first component of B results from the fact that for certain values of X among participants there may be no comparison group members, and vice versa – in formal terms the supports (regions of X where the density function is not zero) of X in the participant and comparison groups may not completely overlap. The second component results from differences in the distribution of X between participants and comparison group members within the region of common support; i.e., for those values of X common to the two groups. The third component represents selection on unobservables as defined in Section 7. This decomposition is helpful for understanding the sources of selection bias as it is conventionally measured.

To reduce the set of conditioning variables, X, down to manageable size, Heckman et al. (1996b, 1998b) condition on the probability of program participation, $P(X)$, rather than directly on X. This is always possible, because we may write the outcome in the absence of training for the experimental controls as follows:

$$Y_0 = \mathrm{E}(Y_0 \mid P(X), D = 1) + V_1,$$

where $\mathrm{E}(V_1 \mid P(X), D = 1) = 0$. The corresponding expression for the comparison group members is given by

$$Y_0 = \mathrm{E}(Y_0 \mid P(X), D = 0) + V_0,$$

where $\mathrm{E}(V_0 \mid P(X), D = 0) = 0$. The residuals average out to zero within participant ($D = 1$) and non-participant ($D = 0$) samples.[82]

Using these methods, this bias B can be decomposed into three components:[83]

$$B = \mathrm{E}(Y_0 \mid D = 1) - \mathrm{E}(Y_0 \mid D = 0) = B_1 + B_2 + B_3. \tag{8.2}$$

To help define B_1, we first define S_P as the common support – the set of $P(X)$ values common to the $D = 1$ and $D = 0$ samples. In addition, let S_{1P} denote the set of $P(X)$ values found in the $D = 1$ sample and S_{0P} the set found in the $D = 0$ sample. The first bias term is given by

$$B_1 = \int_{S_{1P} \cap S_P} \mathrm{E}(Y_0 \mid P(X), D = 1) \mathrm{d}F(P(X) \mid D = 1)$$

[82] This is a valid decomposition whether or not matching is a valid evaluation estimator.

[83] This decomposition was first published in Heckman et al. (1996b).

$$-\int_{S_{0P}\backslash S_P} \mathrm{E}(Y_0 \mid P(X), D=0)\mathrm{d}F(P(X) \mid D=0),$$

where $S_{1P}\backslash S_P$ is the subset of S_{1P} not in S_P, i.e., the set of $P(X)$ values present in the $D=1$ sample but not in the $D=0$ sample. The set $S_{0P}\backslash S_P$ is defined comparably for the $D=0$ group. The second bias term arises from the different densities of $P(X)$ in the $D=1$ and $D=0$ samples:

$$B_2 = \int_{S_P} \mathrm{E}(Y_0 \mid P(X), D=0)[\mathrm{d}F(P(X) \mid D=1) - \mathrm{d}F(P(X) \mid D=0)].$$

The third bias term is the contribution of selection bias rigorously defined:

$$B_3 = P_X \bar{B}_{S_P},$$

where

$$\bar{B}_{S_P} = \frac{\int_{S_P} B(P(X))dF(P(X) \mid D=1)}{\int_{S_P} dF(P(X) \mid D=1)},$$

is the average selection bias defined over the common support set, S_P, and $B(P(X)) = \mathrm{E}(U_0 \mid P(X), D=1) - \mathrm{E}(U_0 \mid P(X), D=0)$ is the selection bias at each point.

The first term on the right-hand side of (8.2) is the difference between the mean earnings of the controls and the comparison group members in the region outside the common support – that is, for those values of $P(X)$ that appear only among controls or only among comparison group members. This is the bias that arises from comparing non-comparable people – persons in $D=1$ who have no counterpart in $D=0$ and vice versa. The second term gives the bias due to the different densities of $P(X)$ in the control and comparison groups over the region in which the densities of $P(X)$ for the two groups overlap. This is the bias that arises from weighting comparable people incomparably.

Finally, the third term, or the "true" selection bias, is the weighted (by the distribution of $P(X)$ for controls) average difference between the earnings of controls and comparisons who have the same $P(X)$. If matching is an effective evaluation method, the third term, B_3, representing selection on unobservables, should be zero or close to it. Recall from the discussion in Section 7.4.1 that under the assumptions that justify matching, $B(P(X))=0$ for all $P(X)$. We can interpret estimates of this term as a measure of the extent to which matching does not balance the bias between treatment and comparison group members.

Heckman et al. (1996b, 1998b) estimate the components of selection bias using the experimental controls from the NJS and a sample of eligible non-participants (ENPs) from the same sites as well as using other, more traditional comparison groups of the sort discussed in Section 7.2.[84] Fig. 10A plots the densities of $P(X)$ for adult male controls

[84] Heckman et al. (1996b, 1998b) estimate the $\mathrm{E}(Y_0 \mid P(X), D=0)$ terms using a local linear regression of the outcome Y_0 on $P(X)$. The estimates of $P(X)$ are obtained from logit models of participation in the JTPA program, but estimates using non-parametric P are very similar.

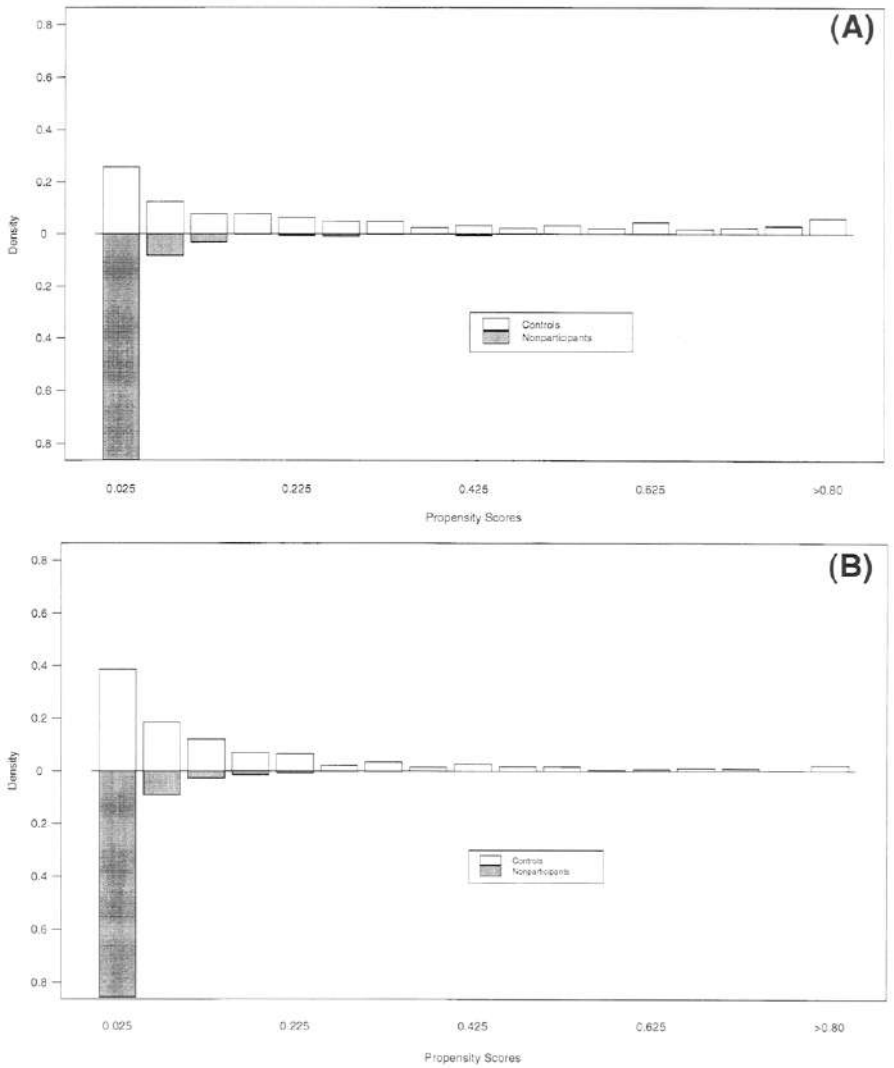

Fig. 10. Density of estimated probability of program particiaption for adult male (A) and female (B) controls and eligible non-participants in the National JTPA Study.

and ENPs. Fig. 10B plots the densities of the $P(X)$ for adult female controls and ENPs. In both groups, for a substantial range of $P(X)$ values in the control sample, there are few or no corresponding comparison group members. Among the adult males, nearly one half of the controls' $P(X)$ values are outside the region of overlapping support.

Table 8 presents estimates of the decomposition in (8.2) for adult males and females in the NJS. As shown by the second row of Table 8, differences in the support of $P(X)$ are an important source of bias. This source of bias is of at least the same order of magnitude as the conventional measure of selection bias presented in the first row of the table. The third row of Table 8 indicates that differences in the distributions of $P(X)$ between control and comparison group members in the region of common support are an important source of bias. Finally the fourth, fifth and sixth rows of the table show that for both groups the selection bias term, B_3, is relatively small compared to the other components of B, the bias as conventionally measured. However, B_3 is still quite large compared to the estimated program impact. This result indicates that matching on $P(X)$ mitigates but does not eliminate selection bias in the NJS data. Selection on unobservables is a substantial component of the experimentally estimated impact of treatment even using the rich data available in the NJS. It is likely to be even more important in cruder datasets, as we document below.

Eliminating selection bias in most non-experimental evaluations may be even more difficult than is suggested by Table 8. The NJS eligible non-participant comparison group was constructed specifically for the purpose of conducting a high quality non-experimental

Table 8
Decomposition of differences in mean earnings for adult participants in the US National JTPA Study (mean monthly earnings differences between experimental controls and comparison sample of eligible non-participants during the 18 months following the baseline in four sites)[a]

	Adult males	Adult females
Mean difference in earnings = B	−337	33
	(47)[b]	(26)
Non-overlapping support = B_1	298	106
	(35)	(13)
	[−88][c]	[318]
Different density weighting of propensity scores = B_2	−659	−118
	(42)	(20)
	[195]	[−355]
Selection bias = B_3	24	45
	(28)	(26)
	[−7]	[136]
Average selection bias when matching only in regions of common support	48	59
Selection bias as a percent of treatment impact	109	202
Control group sample size	508	696
Comparison group sample size	388	866

[a] Source: Heckman et al. (1996b, Table 1, p. 13418).

[b] The numbers in parentheses are the bootstraped standard errors. They are based on 50 replications with 100% sampling.

[c] The numbers in square brackets are the percentage of the mean difference in earnings (row 1) attributable to each component of the bias.

evaluation of JTPA. These data contain many more demographic and baseline characteristics than are commonly available to program evaluators. Further, the comparison group members reside in the same labor market as the trainees, are administered the same survey instruments, and are all eligible for JTPA. The encouraging news from the analyses of Heckman et al. (1997a, 1998b,c) is that less expensive comparison groups that contain limited labor force status histories but still place comparison group members in the same local labor markets as participants and administer the same surveys to both groups should do just as well as the richer data.

Table 9 presents the decomposition when no-shows are used as a comparison group. In the context of the NJS, no-shows are persons randomly assigned to the experimental treatment group who never enroll in JTPA and do not receive JTPA services (these are the dropouts of Section 5). In the absence of an experiment, no-shows are usually persons who enroll in a program but drop out prior to service receipt. Cooley et al. (1979) and Bell et al. (1995) advocate the use of no-shows as a comparison group. On a priori grounds, no-

Table 9
Decomposition of differences in mean earnings in the US National JTPA Study (mean monthly earnings differences during the 18 months following the baseline in four sites, no-shows)[a]

	Experimental controls and treatment group dropouts[b]		Experimental controls and SIPP eligibles[c]	
	Adult males	Adult females	Adult males	Adult females
Mean difference in earnings = B	29	9	−145	47
	(38)[d]	(23)	(56)	(23)
Non-overlapping support = B_1	−13	1	151	97
	(12)	(6)	(30)	(19)
	[−45][e]	[9]	[−104]	[206]
Different density weighting of propensity scores = B_2	3	−9	−417	−172
	(16)	(10)	(44)	(16)
	[11]	[−99]	[287]	[−367]
Selection bias = B_3	38	18	121	122
	(37)	(26)	(33)	(15)
	[135]	[190]	[−83]	[260]
Average selection bias when matching only in regions of common support	42	20	192	198
	(40)	(29)	(57)	(26)
Selection bias as a percent of treatment impact	97	68	440	676

[a] Source: Heckman et al. (1997a, Table 2).

[b] Treatment group dropouts (or "no-shows") are persons randomly assigned to the experimental treatment group who failed to enroll in JTPA.

[c] The SIPP eligibles are persons in the 1998 SIPP full panel who were eligible in month 12 of the 24 month panel using eligibility definition "B" from Devine and Heckman (1996).

[d] Bootstrap standard errors appear in parentheses. They are based on 50 replications with 100% sampling.

[e] The numbers in square brackets are the percentage of the mean difference in earnings (row 1) attributable to each component of the bias.

shows are not necessarily an attractive comparison group. Selective differences in unobservables between participants and no-shows will make the latter a poor comparison group if selection on unobservables (conditional on applying to and being accepted into the program) is an important component of bias. Yet, at the same time, no-shows are an attractive comparison group because they are located in the same labor market and administered the same questionnaire as participants.

The first two columns of Table 9 present the decomposition in (8.2) constructed using the experimental controls and the no-shows from the NJS. Fig. 11A,B presents the densities of $P(X)$ for the same groups. There is much more overlap in the supports of the no-show and control groups than there is in the comparison and control groups. Moreover, the shapes of the distributions of P are closer for no shows and control group members than they are for comparisons and controls (cf. Fig. 10A,B with Fig. 11A,B, respectively).

The evidence on no-shows is mixed. The raw measure of bias B is small for both males and females. In addition, the support and density weighting problems are much smaller than those reported in Table 8, although part of this difference results from the smaller set of X's available in the NJS data to construct $P(X)$ for the no-shows. However, as shown in the final row of Table 9, the selection bias for the no-shows remains sizeable when measured as a percentage of the treatment impact.

The biases obtained for the no-shows in the NJS or the comparison group are much smaller than the biases that result from comparing the NJS controls to a comparison group constructed from a general survey dataset. The last two columns of Table 9 present the bias decompositions based on a comparison group of persons eligible for JTPA drawn from the US Survey of Income and Program Participation (SIPP). The SIPP is a national survey dataset of the type widely used in evaluating active labor market policies. SIPP data are rich enough to determine program eligibility. The comparison group constructed from it is not drawn from the same local labor markets as the NJS control group due to sample size and confidentiality limitations. Moreover, the earnings measure in the SIPP differs substantially from that used for the NJS controls due to differences in the respective survey instruments (Smith, 1997a,b).

A comparison of the first rows of Tables 8 and 9 shows that for the SIPP eligible comparison group, the raw bias, B, is actually smaller for adult males than with the ENP comparison group. The raw bias is about the same magnitude for adult females using the two comparison groups, although of a different sign. However, B_3, selection bias rigorously defined, is much larger for the SIPP eligible comparison group than for the eligible non-participant comparison group in Table 8. This indicates that mismatch of labor markets and questionnaires between participants and comparison group members is a major source of selection bias.

Heckman et al. (1998b) examine these issues in greater depth. In particular, using the NJS data on controls and ENPs, they match controls at two sites with ENPs at the two remaining sites. This comparison shows the effect of putting comparison group members in different local labor markets while holding constant the survey instrument used to measure earnings in the two groups. They find that mismatching the local labor markets

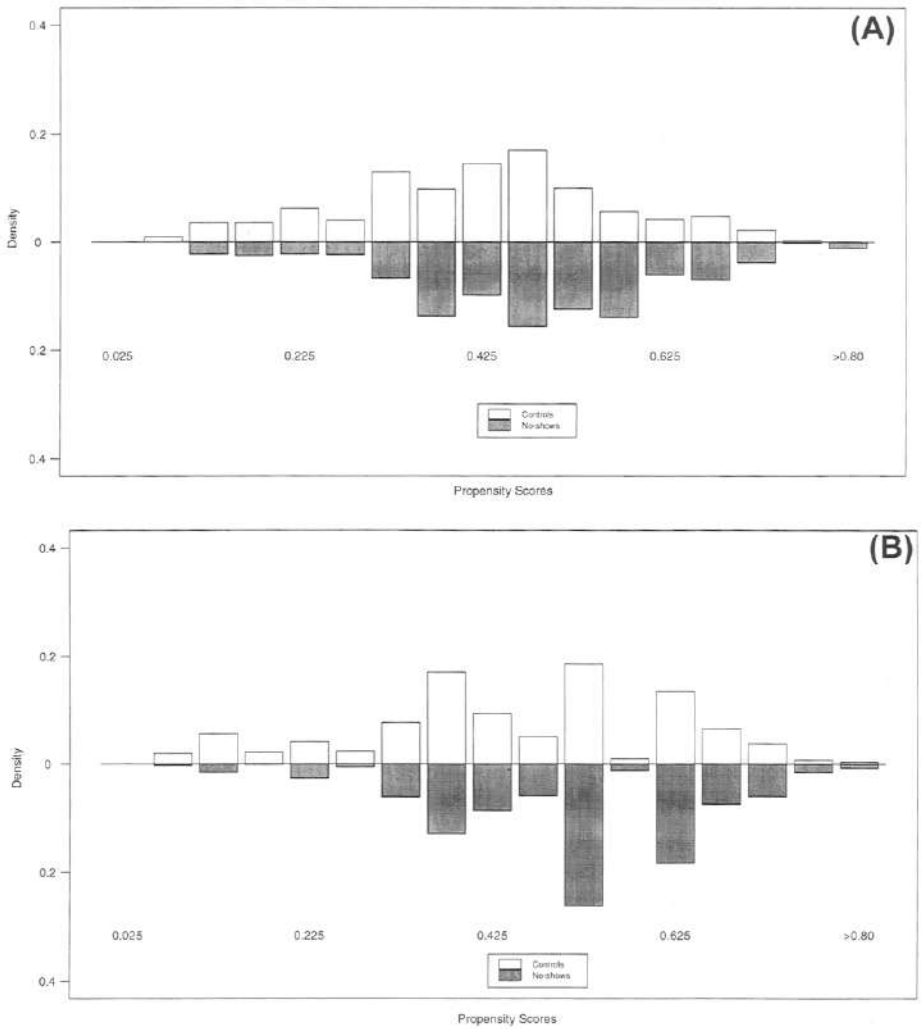

Fig. 11. Density of probability of program participation for adult male (A) and female (B) controls and no-shows in the National JTPA Study.

creates a substantial bias on the order of 30–40% of the estimated treatment effect.[85] Overall, comparing the fifth rows of Tables 8 and 9 suggests that putting participants and comparison group members in the same labor markets and giving them the same

[85] Friedlander and Robins (1995) report similar findings regarding the importance of drawing participants and non-participants from the same local labor markets.

questionnaire eliminates a substantial amount (around 50%) of selection bias, rigorously defined.

Those authors also report that a substantial bias results from using only those observations that fall into the common support of $P(X)$, S_P, for the control and comparison group samples to estimate the impact of treatment. Estimating the experimental treatment effect on the common support rather than on the full support of $P(X)$ among the controls increases the estimate by 50%. Put differently, the experimental impact estimate is higher for persons whose $P(X)$ lies in the common support.

The failure of the common support condition due to an absence of comparison group members comparable to participants in terms of X (or $P(X)$) is a major source of bias in conducting non-experimental evaluations. This motivates one of our major recommendations presented in Section 11 – that non-experimental comparison groups should be designed so that they have the same set of X or $P(X)$ values present among program participants.

An important advantage of an experimental control group in program evaluations is that randomization ensures that the support of treatment and control observed characteristics is the same, up to sampling variation. The results just discussed indicate that non-experimental methods may be able to mitigate major sources of selection bias that arise in the region of common support. Simple principles of using the same questionnaire, locating participants and comparison group members in the same labor markets, comparing comparable people and weighting comparison group members appropriately go a long way toward reducing the conventional measure of selection bias. However, because a significant source of the bias in non-experimental studies is the failure to find a comparison group for which the support of the observed characteristics largely overlaps that of the participants, such studies can only provide a partial description of the impact of treatment. Estimates obtained only over the region of common support may be a poor guide to the impact for all participants. We suspect that this source of bias is substantial for other programs besides the JTPA program where it has been studied.

Heckman et al. (1998b) use the estimated $B(P(X))$ functions to test among competing identifying assumptions for alternative evaluation estimators using the NJS data. Using a variety of X, they reach the following main conclusions:

(I) They reject the assumption: A_M: $B(P(X)) = 0$ for all X which justifies matching;

(II) They *do not* reject the assumption: A_{SS}: $B(X) = B(P(X))$ which says that the bias can be written as a function of $P(X)$ and which justifies the index sufficient classical sample selection model. However, since the support of $P(X)$ is limited, the method cannot recover $E(Y_1 - Y_0 \mid X, D = 1)$ in their data because of the inability to identify the intercepts in the model. They decisively reject the normal sample selection model in their data.

(III) They do not reject the assumption: A_{DD}: $B_t(P(X)) - B_{t'}(P(X)) = 0$ for $t > k > t'$ which justifies the non-parametric difference-in-differences estimator introduced in Heckman et al. (1997a, 1998b). This estimator does not require the full support conditions required in the sample selection estimator although if they are not satisfied, the treatment effect defined only over a subset of the support of $P(X)$.

Finally, even though the assumptions justifying matching are rejected, matching, nonparametric difference-in-differences, and sample selection models do about equally well for the *average* of $E(Y_1 - Y_0 \mid X, D = 1)$ over the support where it can be defined although matching is somewhat inferior to the other two estimators. Their analysis demonstrates that over intervals where the bias balances out, fundamentally different estimators based on different identifying assumptions can identify the same parameter.

Heckman et al. (1998b) emphasize the importance of using semiparametric and nonparametric versions of all three estimators (matching, classical sample selection and difference-in-differences). When they use conventional parametric versions of these estimators, they estimate substantial biases.

The evidence presented in this subsection has major implications for the correct interpretation of LaLonde's (1986) influential examination of the effectiveness of non-experimental evaluation strategies for training programs. As noted in Table 6, LaLonde's non-experimental comparison groups were constructed from various non-comparable data sources. The comparison groups were located in different labor markets from program participants and had their earnings measured in different ways than the participants. His measure of selection bias, B, combines the three factors disentangled in the analyses of Heckman et al. (1996b, 1998b) just summarized.[86] In addition, like most of the studies summarized in Tables 5 and 6, he lacked information on recent preprogram labor force status dynamics which, as noted in Section 6.3, are an important predictor of participation in training. A major conclusion of the analysis of Heckman et al. (1998b) is that a substantial portion of the bias and sensitivity reported by LaLonde is due to his failure to compare comparable people and to weight them appropriately. Further, mismatch of labor markets and questionnaires are also likely important sources of the selection bias measured in LaLonde's study. Overall, the available evidence indicates that simple parametric econometric models applied to bad data do not eliminate selection bias. Instead, better data, including a rich array of X variables for use in constructing $P(X)$, and more appropriate comparison groups, go a long way toward eliminating the sensitivity problems raised in LaLonde's (1986) study.

8.3. A simulation study of the sensitivity of non-experimental methods

A theme of this chapter is that *every* estimator relies on identifying assumptions about the outcome and participation processes. When a particular estimator is applied to data where those assumptions fail to hold, bias results. This bias can be substantial. When different estimators are applied to the same data, the estimates they produce will vary because at most one set of underlying assumptions is consistent with the data. Only if there is no problem of selection bias would all estimators identify the same parameter.

[86] Some of LaLonde's (1986) measures of B are based on a linear regression model that "partials out" X in the sense that linear regression conditions on X. Heckman and Todd (1994) present the appropriate decomposition for this case. When estimated using the NJS controls and eligible non-participants, the same qualitative conclusions emerge about the importance of various components of bias.

To demonstrate these points, in this section we present a simulation study in which we examine the effects of alternative specifications of the processes that determine earnings and participation in training on the performance of various econometric estimators. Using earnings equations and participation rules that are consistent with the evidence from actual training programs, we apply a number of conventional econometric estimators to the simulated data. We vary aspects of the data generating process to see how the different components of the earnings and outcome equations affect the bias of the estimators discussed in Section 7.

8.3.1. A model of earnings and program participation

Building on the model of participation and earnings presented in Section 6.3, we specify a model to underlie our simulation study. Following the notation in Section 6.3, but augmenting it with "i" subscripts to distinguish individual variables from constants, we define the training period as period k, and let D_i be a dummy variable equal to 1 in periods $t > k$ if the individual receives training and 0 otherwise. Prior to the training period ($t < k$), D_i is identically equal to 0 for both future trainees and non-trainees. We further assume that individual i's earnings are determined by the following equation, where the error term combines an AR(1) (autoregressive of order one) process, as used, for example, in Eq. (7.1), with an individual-specific fixed effect, so that

$$Y_{it} = \beta + \alpha_i D_i + \theta_i + U_{it}, \tag{8.3}$$

where

$$U_{it} = \rho U_{i,t-1} + \varepsilon_{it}, \tag{8.4}$$

for all time periods t. $\mathrm{E}(\varepsilon_{it}) = 0$, where ε_{it} is independent and identically distributed over time and persons. The individual-specific fixed effect, θ_i, is drawn from a population distribution with mean zero. We assume that $\theta_i, \varepsilon_{it}$ and α_i are mutually independent. We assume random sampling so that all i-subscripted random variables are statistically independent of all i' subscripted variables, $i \neq i'$.

In this model,

$$Y_{it} = D_i Y_{1it} + (1 - D_i) Y_{0it}$$

and

$$Y_{1it} - Y_{0it} = \alpha_i.$$

This is a random coefficients model in which the effect of training, α_i, varies among individuals according to some population distribution. This specification of the outcome equation yields two parameters of interest: the mean effect of training in the population, $\mathrm{E}(\alpha_i)$, and the mean effect of training on those who actually receive training, $\mathrm{E}(\alpha_i \mid D_i = 1)$. The more standard common coefficient specification assumes that $\alpha_i = \alpha$ for all individuals, in which case the two parameters are equivalent.

Following the model of perfect certainty presented in Section 6.3, we assume that the

decision to participate in training depends on individuals' discounted lifetime gain from training, α_i/r, their opportunity costs or foregone earnings in period k, Y_{ik}, and their tuition costs or subsidy, c_i. More formally, we have

$$D_i = \begin{cases} 1 & \text{if } \alpha_i/r - Y_{ik} - c_i > 0 \text{ and } t > k, \\ 0 & \text{otherwise.} \end{cases} \tag{8.5}$$

As noted in Section 6.3, this model is consistent with Ashenfelter's dip in earnings among participants prior to participation. In some of the specifications analyzed below, we relax the perfect foresight assumption and consider the case where α is not known by the agent at the time program participation decisions are made. In Eq. (8.5), we introduce instruments as determinants of program costs and write $c_i = Z_i'\phi + V_i$, where Z_i is an observed characteristic that affects the cost of training and where V_i is a mean zero random disturbance. For simplicity, we assume that both Z_i and V_i are independent of all other variables and errors. We assume the trainees have zero earnings during the training period. Because D_i depends on foregone or "latent" earnings in period k, $E(D_i\theta_i)$ is non-zero and, in fact, is negative. Persons with higher values of θ_i have higher opportunity costs. As a result, OLS estimates of our parameters of interest are downward biased.

8.3.2. The data generating process

In our simulations, we set β=1000 and the treatment effect, α_i, is drawn from a normal distribution with a mean of 100 and standard deviation of σ_α. We explore the effects on the bias of different values of σ_α, including the common effect model where $\sigma_\alpha = 0$. The ε_{it} are randomly drawn from a normal distribution with mean zero and standard deviation σ_ε. We initialize the process by setting $U_{i,k-5} = \varepsilon_{i,k-5}$, where $k-5$ is the initial period in the simulated data. We generate the θ_i from a normal distribution with mean zero and standard deviation σ_θ.

In the participation equation, the Z_i are randomly drawn from a $N(\mu_Z,\sigma_Z^2)$ distribution and the parameter ϕ is set equal to 1. The mean of the distribution of characteristics, μ_Z, is chosen so that, for each simulated sample, 10% of the population enters the program. Notice that because we draw the characteristics, Z_i, independently of both components of the outcome equation unobservable, θ_i and ε_{it}, Z_i is a valid instrument for the training variable D_i in the common coefficient model. When α_i varies among persons, and is acted on by agents, Z_i is not a valid instrument for the parameter $E(\alpha_i \mid D_i = 1)$ for the reasons given in Section 7.4.3. Only if the idiosyncratic component of α_i is not acted on in making participation decisions is IV a valid estimator of $E(\alpha_i \mid D_i = 1)$. We set the discount rate r to be 0.10. To complete the parameterization of the participation equation, we draw the disturbances, V_i, from a $N(0,\sigma_V^2)$ distribution.

Using this specification, in most of the runs we generate 100 samples each containing 1000 individuals. For each person in each sample, we generate 10 periods of earnings data. There are five pre-program periods, $k-5$ to $k-1$, one training period, k, and four post-program periods, $k+1$ to $k+4$ that we simulate. However, persons are assumed to live forever so the simple infinite horizon decision rule applies. Each sample consists of 100

participants and 900 non-participants. The "unmatched" comparison group used in the tables consist of all of the non-participants. Tables 10, 12 and 13 present estimates using unmatched comparison groups.

Matched samples are often formed prior to applying econometric estimators. As noted in Section 7.2, applying estimators to matched samples often invalidates the properties of an estimator that is appropriate in random or unmatched samples. In fact, matching is an estimator in its own right. The conventional practice of matching and then using an econometric estimator on the new samples created by matching is not in general justified. To illustrate the effects of this practice, our matched comparison group consists of non-participants matched to the participants using nearest neighbor matching with replacement. The sample sizes for the matched samples are much smaller. We have 100 treatment group members as before but at most 100 unique comparison group members in each matched sample – compared to the 900 members in the unmatched comparison group. Unless otherwise stated, the matching is on earnings two periods prior to participation, i.e., on $Y_{i,k-2}$. Similar matching or screening rules are widely used in the literature. Tables 11, 14 and 15 present estimates using the matched comparison groups, with the latter two tables examining the effects of alternative matching rules.

In the first column of Table 10, we present a set of "base case" estimates for a variety of models with a data generating process $\theta_i \sim N(0, 300)$, $\varepsilon_i \sim N(0, 450)$, $Z_i \sim N(0, 300)$, $\rho = 0.78$, and $\alpha_i = 100 + N(0, 300)$. These distributions are chosen to represent samples of the sort that appear in practice. The values for the standard deviations of θ_i and ε_i, as well as the value of ρ, are based on estimates reported in Ashenfelter and Card (1985). The value for the standard deviation of α_i is based on the estimate reported in Heckman et al. (1997c).[87] Column (1) considers the base case when $\mathrm{E}(\alpha_i \mid D_i = 1)$ is the parameter of interest while column (3) considers the base case when $\mathrm{E}(\alpha_i)$ is the parameter of interest. The expected value of the parameter of interest taken over all 100 simulated datasets appears in the column heading for each specification. In the base case, $\mathrm{E}(\alpha_i \mid D_i = 1) = 607.8$. Given that $\mathrm{E}(\alpha_i) = 100$, this indicates substantial selection into the program based on α_i. As previously discussed, the bias for $\mathrm{E}(\alpha_i)$ is the bias for $\mathrm{E}(\alpha_i \mid D_i = 1)$ plus $\mathrm{E}(\alpha_i - \mathrm{E}(\alpha_i) \mid D = 1) = \mathrm{E}(U_{1i} - U_{0i} \mid D = 1)$, the term incorporated into the definition of $\mathrm{E}(\alpha_i \mid D_i = 1)$.

In the remaining columns of Table 10, we vary one aspect of the data generating process at a time using the base case as a point of departure. Column (2) presents the common coefficient case, with $\alpha_i = \alpha = 100$ for all i. Column (4) presents the case of a random coefficient model where agents know $\mathrm{E}(\alpha_i)$ rather than α_i when making their program participation decisions. Thus there is ex ante homogeneity but ex post heterogeneity in realized outcomes so $\mathrm{E}(\alpha_i \mid D_i = 1) = \mathrm{E}(\alpha_i)$ and Z_i is a valid instrument for both parameters. Column (5) presents the base case with an increased variance of α_i. For each

[87] If α_i is a log concave random variable, then in a Roy model, the Heckman et al. (1997c) estimates of the variance of α_i are understated since they estimate $\mathrm{Var}(\alpha_i \mid D_i = 1)$ and not $\mathrm{Var}(\alpha_i)$.

specification in Table 10, Table 11 presents estimates using the matched comparison group sample.

In Section 7 we focused primarily on identification of the various parameters of interest under different assumptions about data generating processes. This focus follows much of the recent econometric literature on program evaluation, starting with Heckman and Robb (1985a, 1986a). In practice, securing identification is only a useful first step in determining a valid estimation strategy. The sampling variability of alternative estimators is an important consideration in picking an estimator. Different estimators converge to the true value at different rates. Table 12 presents some Monte Carlo evidence on the rates of convergence of the estimators we examine using different sample sizes.

Table 13 presents the results from simulations in which we reduce the standard deviations of the random variables determining outcomes and participation one at a time, holding the overall variances fixed, in order to explore the effect of the size of relative components of variance on the bias.

We stress that the Monte Carlo analysis reported in this chapter is illustrative rather than definitive. Heckman and Smith (1998e) present a much more comprehensive Monte Carlo study which examines the bias and small sample variability of the main non-experimental estimators presented in Section 7. Our work draws from their findings.

8.3.3. The estimators we examine

The assumptions required to justify each estimator are discussed in Sections 3, 4 and 7. Here we briefly discuss how each estimator was implemented in our simulation study. The estimators selected are those most commonly used in the literature. The entries in the tables indicate the mean and, in parentheses, the standard deviation of the estimates obtained from the 100 simulated samples. For the IV and Heckman (1979) estimators we present additional statistics of interest. Unless otherwise noted, the estimates presented in these tables reflect impacts on Y_{k+4}.

The first estimator in each table is the cross-section estimator applied to post-program earnings. Because we do not include any observables in the earnings equation, the cross-section estimator is the coefficient on D_i in a regression of $Y_{i,k+4}$ on D_i and is equivalent to the difference between the mean of participant and non-participant earnings. The cross-section estimator is biased downward when $\mathrm{Var}(\theta_i) > 0$ or $\rho > 0$. When $\mathrm{Var}(\theta_i) = \rho = 0$, the cross-section estimator identifies $\mathrm{E}(\alpha_i \mid D_i = 1)$ when applied to post-program earnings.

The second, third and fourth rows in each table present three alternative versions of the difference-in-differences estimator based on the averages (over $D = 1$ and $D = 0$) of the comparisons:

$$Y_{it} - Y_{it'} = \alpha_i D_i + (U_{it} - U_{it'}), \tag{8.6}$$

where $t' < k < t$. In all three rows, $t = k + 3$ is the "after" period. The three rows differ based on the value chosen for the "before" period, to show the effect of differencing relative to different points in the sequence along Ashenfelter's dip and also to illustrate the symmetric differencing estimator. In the second row, the before period is $k - 1$, in the

Table 10
Bias in non-experimental estimates of the impact of training (unmatched comparison group samples)[a]

Estimator[b]	Base case[c]; parameter of of interest: $E(\alpha \mid D = 1) = 615.7$ (1)	Base case with common coefficient; parameter of interest: $E(\alpha \mid D = 1) = 100.0$ (2)	Base case; parameter of interest: $E(\alpha) = 100.0$ (3)	Base case where agent knows $E(\alpha)$, not α_i; parameter of interest $E(\alpha \mid D = 1) = 98.4$ (4)	Base case with increased variance of α; parameter of interest $E(\alpha \mid D = 1) = 971.4$ (5)
Cross-section					
Mean	−98.5	−494.7	417.2	−494.7	−60.5
SD	(61.9)	(47.9)	(66.6)	(47.9)	(59.2)
Diff-in-diff (−1,3)					
Mean	34.0	155.6	549.7	155.6	18.7
SD	(58.0)	(54.4)	(61.4)	(54.4)	(56.1)
Diff-in-diff (−3,3)					
Mean	−9.8	−55.2	505.9	−55.2	−7.9
SD	(60.7)	(52.3)	(64.5)	(52.3)	(58.5)
Diff-in-diff (−5,3)					
Mean	−40.2	−211.7	475.5	−211.7	−26.9
SD	(63.1)	(51.6)	(67.1)	(51.6)	(57.9)
AR(1) regression					
Mean	−15.7	−187.1	500.0	−174.0	5.0
SD	(203.5)	(203.3)	(202.9)	(210.3)	(197.4)
IV estimator					
Mean	342.7	−15.3	858.4	−15.1	−102.8
Median	−146.0	−5.8	369.1	−5.7	−247.7
SD	(4829.0)	(231.5)	(4829.1)	(227.7)	(5131.2)
Corr(Z,D)	0.0559	0.2755	0.0559	0.2755	0.0355

Ashenfelter (1979)					
Mean	75.9	166.4	591.6	166.1	86.5
SD	(59.3)	(53.3)	(62.7)	(53.7)	(58.8)
Heckman (1979)					
Mean	214.5	53.2	1139.8	54.8	350.2
Median	−90.9	53.5	204.8	61.2	−178.3
SD	(3595.3)	(221.4)	(12811.0)	(269.1)	(6618.0)
Kitchen sink					
Mean	−20.0	−160.7	495.7	−160.9	−11.9
SD	(54.9)	(54.8)	(58.6)	(55.0)	(52.4)

[a] Estimates are based on 100 simulated samples of 1000 observations each. The "mean" row presents the mean of the estimates from the 100 samples while the "SD" row presents the standard deviation of the estimates from the 100 samples. The "Corr(Z,D)" row for the IV estimates gives the average correlation between the participation indicator, D, and the instrument, Z.

[b] The cross-section estimator is the simple difference between participant and non-participant earnings in period $k + 4$. The difference-in-differences estimates are based on the periods indicated, so that $(-1,3)$ is the difference between the change in participant earnings from period $k-1$ to period $k + 3$ and the change in non-participant earnings over the same interval. The difference-in-differences $(-3,3)$ estimator is symmetric. The AR(1) estimates are based on a regression of Y_{k+4} on Y_{k+3} and D, with the estimate consisting of the coefficient estimate on D divided by $(1-\rho)$, where ρ is estimated by the coefficient on Y_{k+3}. The IV estimates use Z as an instrument for a regression of Y_{k-4} on D. The Ashenfelter (1979) estimator is described in Section 8.3.3. The dependent variable for this estimator is $Y_{i+4} - Y_k$. The Heckman (1979) estimator is a special case of the class of control function estimators presented in Section 7.4.2. In columns (1) and (5) the estimate is calculated as shown in Section 7.4.2. In columns (2), (3) and (4) the estimate is the coefficient on D when the estimated control functions are included. The dependent variable for the Heckman (1979) estimator is Y_{k+4}. The kitchen sink estimates are based on a regression of Y_{k+4} on Y_{k-1}, Y_{k-2} and Z.

[c] The base case has $\theta \sim N(0,300)$, $\varepsilon \sim N(0,280)$, $Z \sim N(0,300)$, $V \sim N(0,200)$, $\rho = 0.78$ and $\alpha = 100 + N(0,300)$. This case is based on estimates of the size of the permanent and transitory components of earnings from Ashenfelter and Card (1985) and of the variance in the impacts of training from Heckman et al. (1997c). In column (2), $\alpha = 100$. In column (5), $\alpha = 100 + N(0,500)$. In the base case in columns (1) and (3), the fractions of $\text{Var}(Y_{k+4} \mid D = 1)$ accounted for by α and θ are 0.0564 and 0.2670, respectively. In column (2), they are 0.0000 and 0.3132, respectively. In column (4), they are 0.2787 and 0.2246, respectively. In column (5), they are 0.1273 and 0.2467, respectively.

Table 11
Bias in non-experimental estimates of the impact of training (matched comparison group samples)[a]

Estimator[b]	Base case[c]; parameter of interest $E(\alpha\|D=1)=615.7$ (1)	Base case with common coefficient; parameter of interest $E(\alpha\|D=1)=100.0$ (2)	Base case; parameter of interest $E(\alpha)=100.0$ (3)	Base case where agent knows $E(\alpha)$, not α_i; parameter of interest $E(\alpha\|D=1)=98.4$ (4)	Base case with increased variance of α; parameter of interest $E(\alpha\|D=1)=971.4$ (5)
Cross-section[c]					
Mean	−42.9	−233.0	472.8	−233.0	−27.4
SD	(80.3)	(70.4)	(81.8)	(70.4)	(74.4)
Diff-in-diff (−1,3)					
Mean	−5.8	−36.8	509.9	−36.8	−5.9
SD	(77.5)	(77.1)	(78.8)	(77.1)	(77.5)
Diff-in-diff (−3,3)					
Mean	−43.3	−243.2	472.4	−243.2	−28.1
SD	(82.5)	(70.9)	(84.4)	(70.9)	(82.2)
Diff-in-diff (−5,3)					
Mean	−33.9	−202.0	481.8	−202.0	−23.2
SD	(80.3)	(76.1)	(83.0)	(76.1)	(77.9)
AR(1) regression					
Mean	−6.9	−69.1	508.8	−13.7	5.4
SD	(333.7)	(479.3)	(333.7)	(588.9)	(323.3)
IV estimator					
Mean	−305.0	27.3	210.7	28.2	−427.5
Median	−41.9	18.1	474.1	10.1	36.2
SD	(4001.6)	(175.8)	(4000.8)	(191.0)	(3643.9)
Corr(Z,D)	0.0923	0.5035	0.0923	0.5035	0.0604
Ashenfelter (1979)					
Mean	81.5	209.3	597.2	208.6	90.0
SD	(83.0)	(85.0)	(84.7)	(84.7)	(79.3)

Heckman (1979)					
Mean	−22060.2	24.1	−6915.5	22.9	221.7
Median	−57.6	13.9	469.0	7.3	−41.5
SD	(222223.4)	(177.2)	(71920.5)	(198.5)	(3739.7)
Kitchen sink					
Mean	−22.8	−194.7	492.9	−194.5	−13.9
SD	(80.2)	(91.5)	(81.1)	(94.9)	(75.4)

[a] Estimates are based on 100 simulated samples of 1000 observations each. The "mean" row presents the mean of the estimates from the 100 samples while the "SD" row presents the standard deviation of the estimates from the 100 samples. The "Corr(Z,D)" row for the IV estimates gives the average correlation between the participation indicator, D, and the instrument, Z. Matching consists of nearest neighbor matching on Y_{k-2} with replacement. The average number of unique observations in a matched sample is 92.1 in columns (1) and (3), 79.8 in columns (2) and (4) and 92.3 in column (5).

[b] The cross-section estimator is the simple difference between participant and non-participant earnings in period $k+4$. The difference-in-differences estimates are based on the periods indicated, so that (−1,3) is the difference between the change in participant earnings from period $k-1$ to period $k+3$ and the change in non-participant earnings over the same interval. The difference-in-differences (−3,3) estimator is symmetric. The AR(1) estimates are based on a regression of Y_{k+4} on Y_{k-3} and D, with the estimate consisting of the coefficient estimate on D divided by $(1-\rho)$, where ρ is estimated by the coefficient on Y_{k-3}. The IV estimates use Z as an instrument for a regression of Y_{k+4} on D. The Ashenfelter (1979) estimator is described in Section 8.3.3. The dependent variable for this estimator is $Y_{k+4}-Y_k$. The Heckman (1979) estimator is a special case of the class of control function estimators presented in Section 7.4.2. In columns (1) and (5) the estimate is calculated as shown in Section 7.4.2. In columns (2), (3) and (4) the estimate is the coefficient on D when the estimated control functions are included. The dependent variable for the Heckman (1979) estimator is Y_{k+4}. The kitchen sink estimates are based on a regression of Y_{k+4} on Y_{k-1}, Y_{k-2} and Z.

[c] The base case has $\theta \sim N(0,300)$, $\varepsilon \sim N(0,280)$, $Z \sim N(0,300)$, $V \sim N(0,200)$, $\rho = 0.78$ and $\alpha = 100 + N(0,300)$. This case is based on estimates of the size of the permanent and transitory components of earnings from Ashenfelter and Card (1985) and of the variance in the impacts of training from Heckman et al. (1997c). In column (2), $\alpha = 100$. In column (5), $\alpha = 100 + N(0,500)$. In the base case in columns (1) and (3), the fractions of $\text{Var}(Y_{k+4} \mid D = 1)$ accounted for by α and θ are 0.0564 and 0.2670, respectively. In column (2), they are 0.0000 and 0.3132, respectively. In column (4), they are 0.2787 and 0.2246, respectively. In column (5), they are 0.1273 and 0.2467, respectively.

Table 12
Bias in non-experimental estimates of the impact of training (unmatched comparison group samples)[a]

Estimator[b]	Base case[c] with sample size = 2500; parameter of interest $E(\alpha \mid D = 1) = 615.1$ (1)	Base case with sample size = 5000; parameter of interest $E(\alpha \mid D = 1) = 614.6$ (2)	Base case with sample size = 10000; parameter of interest $E(\alpha \mid D = 1) = 614.6$ (3)
Cross-section			
Mean	−102.0	−106.0	−103.4
SD	(35.4)	(24.7)	(18.8)
Diff-in-diff (−1,3)			
Mean	34.5	34.5	34.6
SD	(35.1)	(25.3)	(17.0)
Diff-in-diff (−3,3)			
Mean	−13.0	−13.7	−10.5
SD	(38.5)	(26.7)	(18.4)
Diff-in-diff (−5,3)			
Mean	−48.4	−47.3	−44.4
SD	(33.8)	(22.0)	(16.8)
AR(1) regression			
Mean	−4.4	−20.9	−26.5
SD	(137.1)	(85.8)	(65.2)
IV estimator			
Mean	−191.2	−201.1	−173.4
Median	−118.7	−205.1	−170.8
SD	(837.4)	(470.9)	(322.2)
Corr(Z,D)	0.0536	0.0577	0.0577
Ashenfelter (1979)			
Mean	73.0	71.0	70.3
SD	(30.1)	(21.8)	(15.5)
Heckman (1979)			
Mean	2.0	−60.5	−38.8
Median	24.8	−27.6	−20.6
SD	(931.8)	(584.1)	(380.8)
Kitchen sink			
Mean	−20.3	−23.2	−22.6
SD	(30.2)	(20.9)	(15.9)

[a] Estimates are based on 100 simulated samples of the indicated size. The "mean" row presents the mean of the estimates from the 100 samples while the "SD" row presents the standard deviation of the estimates from the 100 samples. The "Corr(Z,D)" row for the IV estimates gives the average correlation between the participation indicator, D, and the instrument, Z.

[b] The "base case" has $\theta \sim N(0, 300)$, $\varepsilon \sim N(0, 280)$, $Z \sim N(0, 300)$, $V \sim N(0, 200)$, $\rho = 0.0$, $\alpha = 100 + N(0, 300)$. Estimates for the base case with samples of size 1000 appear in Table 10. This case is based on estimates of the size of the permanent and transitory components of earnings from Ashenfelter and Card (1985) and of the variance in the impacts of training from Heckman et al. (1997c). In column (1), the fractions of $\mathrm{Var}(Y_{k+4} \mid D = 1)$ accounted for by α and θ are 0.0556 and 0.2678, respectively. In column (2), the fractions are 0.0561 and 0.2692, respectively. In column (3), the fractions are 0.0558 and 0.2695, respectively.

third row it is $k-3$, which is the symmetric case, and in the fourth row it is $k-5$. The general difference-in-differences estimator will only be consistent for $\mathrm{E}(\alpha_i \mid D_i = 1)$ when $\rho = 0$.

The fifth estimator is the simple autoregressive estimator discussed in Section 7.6:

$$Y_{it} = \rho Y_{i,t-1} + (1-\rho)\beta + (1-\rho)\mathrm{E}(\alpha_i \mid D_i = 1)D_i + (1-\rho)\theta_i$$

$$+(1-\rho)D_i[\alpha_i - \mathrm{E}(\alpha_i \mid D_i = 1)] + \varepsilon_{it}$$

$$= \rho Y_{i,t-1} + \beta^* + \alpha^* D_i + \theta^* + (1-\rho)D_i[\alpha_i - \mathrm{E}(\alpha_i \mid D_i = 1)] + \varepsilon_{it}, \tag{8.7}$$

where $\beta^* = (1-\rho)\beta$ and $\alpha^* = (1-\rho)\mathrm{E}(\alpha_i \mid D_i = 1)$. We define $\hat{\alpha}_{\mathrm{AR}} = \hat{\alpha}^*/(1-\hat{\rho})$ where $\hat{\alpha}^*$and $\hat{\rho}$ are the OLS estimators of $(1-\rho)\mathrm{E}(\alpha_i \mid D_i = 1)$ and ρ, respectively. The autoregressive estimator identifies $\mathrm{E}(\alpha_i \mid D_i = 1)$ only when $\mathrm{Var}(\theta_i) = 0$ and $\sigma_\alpha = 0$, i.e., only when there are no fixed effects in the outcome equation and there is no heterogeneity in the impact of treatment.

The sixth estimator we consider is an instrumental variables (IV) estimator. We calculate the IV estimates using Z_i, the observable variable in the participation equation, as an instrument for the training indicator variable, D_i, in earnings Eq. (8.3). For post-program earnings, the IV estimator will consistently estimate $\mathrm{E}(\alpha_i \mid D_i = 1)$ if $\mathrm{E}(Z_i D_i) \neq 0$, $\mathrm{E}(Z_i \theta_i) = 0$, and $\mathrm{E}(Z_i \varepsilon_i) = 0$ for all t and if α_i is the same for everyone or, when it is heterogeneous, if agents do not choose to participate in the program based upon it. If agents select into the program based on α_i, then IV is inconsistent for $\mathrm{E}(\alpha_i \mid D_i = 1)$. However, in this case IV estimates the LATE associated with the instrument Z_i because our model satisfies the monotonicity and independence conditions (7.IA.1) and (7.IA.2) of Imbens and Angrist (1994). Accordingly, provided that the estimates converge adequately to large sample values, our Monte Carlo analysis reveals how much the LATE differs from treatment on the treated assuming that the estimator is consistent.

The seventh estimator we consider is Ashenfelter's (1979) difference-in-differences autoregressive estimator. His estimator may be written as

$$Y_{it} - Y_{ik} = (\rho^{t+1} - \rho)Y_{i,k-1} + \beta^{**} + \alpha^{**} D_{it} + \theta^{**} + U^{**}. \tag{8.8}$$

From knowledge of ρ, Ashenfelter proposes to estimate the parameter of interest, $\mathrm{E}(\alpha_i \mid D_i = 1)$, using

$$\hat{\alpha}_{ASH} = \left[\sum_{j=1}^{t-1} \rho^j (1-\rho)\right]^{-1} \hat{\alpha}^{**}, \tag{8.9}$$

where $\hat{\alpha}^{**}$ is the OLS estimate of α^{**} in Eq. (8.8). When $\rho \neq 0$, this estimator is biased and inconsistent for α in the common coefficient model and for both $\mathrm{E}(\alpha_i)$ and $\mathrm{E}(\alpha_i \mid D_i = 1)$ in the random coefficient model (Heckman, 1978).

The eighth estimator shown in each table is the Heckman (1979) two-step estimator based on the assumption that the unobservables in the outcome and participation equations

Table 13
Bias in non-experimental estimates of the impact of training (unmatched comparison group samples)[a]

Estimator[b]	Base case[c] with reduced variance of θ; parameter of interest $E(\alpha \mid D = 1) = 616.2$ (1)	Base case with reduced variance of ε; parameter of interest $E(\alpha \mid D = 1) = 615.1$ (2)	Base case with reduced variance of z; parameter of interest $E(\alpha \mid D = 1) = 615.8$ (3)	Base case with reduced variance of V; parameter of interest $E(\alpha \mid D = 1) = 616.1$ (4)	Base case with $\rho = 0$ and no fixed effect, θ; parameter of interest $E(\alpha \mid D = 1) = 615.3$ (5)
Cross-section					
Mean	−59.9	−185.4	−96.7	−97.9	7.5
SD	(58.0)	(50.7)	(55.6)	(61.0)	(57.8)
Diff-in-diff (−1.3)					
Mean	46.7	0.4	33.7	32.1	−0.4
SD	(69.2)	(6.2)	(53.3)	(58.7)	(89.2)
Diff-in-diff (−3.3)					
Mean	−15.8	−0.3	−11.9	−9.8	−3.1
SD	(71.7)	(6.2)	(59.9)	(59.9)	(78.0)
Diff-in-diff (−5.3)					
Mean	−61.6	−0.3	−41.9	−40.1	2.4
SD	(72.3)	(5.9)	(58.9)	(60.7)	(91.2)
AR(1) regression					
Mean	25.1	−445.6	−0.7	−20.2	7.4
SD	(157.1)	(4510.8)	(192.2)	(193.7)	(57.7)
IV estimator					
Mean	−605.0	−193.2	−2054.7	119.2	−207.6
Median	−239.1	−323.7	−230.0	−164.7	−147.2
SD	(3630.6)	(2020.2)	(17443.0)	(2678.8)	(1962.8)
Corr(Z,D)	0.0552	0.0576	0.0063	0.0677	0.0583

Ashenfelter (1979)					
Mean	96.8	43.4	76.1	73.5	247.7
SD	(64.7)	(6.6)	(55.5)	(56.6)	(83.6)
Heckman (1979)					
Mean	−887.5	50.0	−2515.9	290.5	−446.7
Median	−129.8	−56.2	−98.5	−21.7	−109.3
SD	(5429.7)	(3310.5)	(19052.7)	(3162.0)	(3546.4)
Kitchen sink					
Mean	−19.8	2.0	−18.9	−21.2	8.7
SD	(59.2)	(6.2)	(48.2)	(53.8)	(58.3)

[a] Estimates are based on 100 simulated samples of 1000 observations each. The "mean" row presents the mean of the estimates from the 100 samples while the "SD" row presents the standard deviation of the estimates from the 100 samples. The "Corr(Z,D)" row for the IV estimates gives the average correlation between the participation indicator, D, and the instrument, Z.

[b] The cross-section estimator is the simple difference between participant and non-participant earnings in period $k+4$. The difference-in-differences estimates are based on the periods indicated, so that $(-1,3)$ is the difference between the change in participant earnings from period $k-1$ to period $k+3$ and the change in non-participant earnings over the same interval. The difference-in-differences $(-3,3)$ estimator is symmetric. The AR(1) estimates are based on a regression of Y_{k+4} on Y_{k+3} and D, with the estimate consisting of the coefficient estimate on D divided by $(1-\rho)$, where ρ is estimated by the coefficient on Y_{k+3}. The IV estimates use Z as an instrument for a regression of Y_{k-4} on D. The Ashenfelter (1979) estimator is described in Section 8.3.3. The dependent variable for this estimator is $Y_{k-4}-Y_k$. The Heckman (1979) estimator is a special case of the class of control function estimators presented in Section 7.4.2. The estimates in all five columns are calculated as shown in Section 7.4.2. The dependent variable for the Heckman (1979) estimator is Y_{k-4}. The kitchen sink estimates are based on a regression of Y_{k+4} on Y_{k-1}, Y_{k-2} and Z.

[c] The base case has $\theta \sim N(0,300)$, $\varepsilon \sim N(0,280)$, $Z \sim N(0,300)$, $V \sim N(0,200)$, $\rho = 0.78$ and $\alpha = 100 + N(0,300)$. In column (1), $\theta \sim N(0,30)$ and $\varepsilon \sim N(0,337)$, in column (2), $\theta \sim N(0,538)$ and $\varepsilon \sim N(0,30)$, in column (3), $Z \sim N(0,30)$ and $V \sim N(0,359)$, in column (4), $Z \sim N(0,359)$ and $V \sim N(0,30)$ and in column (5), $\rho = 0$ and there is no fixed effect, θ. The base case is based on estimates of the size of the permanent and transitory components of earnings from Ashenfelter and Card (1985) and of the variance in the impacts of training from Heckman et al. (1997c). In column (1), the fractions of $\mathrm{Var}(Y_{k+4} \mid D=1)$ accounted for by α and θ are 0.0578 and 0.0028, respectively. In column (2), the fractions are 0.0540 and 0.8015, respectively. In column (3), the fractions are 0.0563 and 0.2683, respectively. In column (4), the fractions are 0.0560 and 0.2670, respectively. In column (5), the fractions are 0.0605 and 0.0000, respectively.

are jointly normally distributed. The general control function estimator, of which the Heckman (1979) estimator is a special case, is given by Eq. (7.15). Under its identifying assumptions, this estimator consistently estimates both $E(\alpha_i)$ and $E(\alpha_i \mid D_i = 1)$ using the procedures described in Section 7.4.2. Because we assume normal errors, our analysis is favorable to this estimator.

The final row in each table presents what we call the "kitchen sink" estimator. This estimator approximates the common practice of conditioning on whatever variables are available in an earnings equation that also includes an indicator for receipt of training. The Barnow et al. (1980) estimator is a version of the kitchen sink estimator (see the discussion in Section 7.4.1). We implement this estimator by regressing earnings in each post-program period on D_i, X_i, $Y_{i,k-2}$, and $Y_{i,k-1}$. This estimator is inconsistent for all of the specifications we consider except those with $\rho = 0$.

8.3.4. Results from the simulations

All of the specifications we consider depart from the base case presented in the first columns of Tables 10 (for the unmatched comparison group) and 11 (for the matched comparison group). In the base case, the cross-section estimator is biased downward in both the unmatched and matched samples because persons with low fixed effects, θ_i, are differentially more likely to participate in the program, which implies that participants have lower average earnings without training than do comparison group members. This bias is accentuated by selection into the program based on low values of ε_{it} in the enrollment period k, which persist over time due to the high value of ρ. Using a matched comparison group cuts the mean bias for the cross-section estimator roughly in half. The difference-in-differences estimator takes care of the selection on θ_i when that is the only source of bias, but not the selection bias due to the persistence in the transitory shocks. It has a lower mean bias than the cross-section estimator but is still inconsistent. Use of a matched comparison group has mixed effects on the bias in the difference-in-differences estimator.

The AR(1) estimator is consistent if $\sigma_\theta = 0$ and $\sigma_\alpha = 0$. In the base case, even though $\sigma_\theta > 0$ and $\sigma_\alpha > 0$, the estimator performs relatively well, with the lowest mean bias for the unmatched comparison group and one of the lowest with the matched comparison group. This is an artifact of the specific parameter values chosen for the base case model. For this model, several sources of bias just happen to cancel out, resulting in a lower overall bias (Heckman and Smith, 1998e). In particular, $Y_{i,t-1}$ is positively correlated with both θ_i and $D_i(\alpha_i - E(\alpha_i \mid D_i = 1))$ in the outcome equation error, and D_i is negatively correlated with θ_i. Heckman and Smith (1998e) present a comprehensive analysis of this case and demonstrate that perturbations in the base case specifications produce large biases in the AR(1) model.

The IV estimator is inconsistent for treatment on the treated in the base case because Z_i is correlated with the error term conditional on D_i as shown in Section 7.4.3. This inconsistency is reflected in large and highly variable biases with both the matched and unmatched comparison groups. However, IV consistently estimates the LATE associated

with Z_i. Using the median value, the LATE parameter is 25% lower than the treatment on the treated parameter. The Ashenfelter (1979) and kitchen sink estimators are inconsistent as well, but have relatively small estimated biases. In both cases, conditioning on lagged earnings appears to provide an imperfect but still helpful control for the effects of selection in both the matched and unmatched samples.

Column (2) of Tables 10 and 11 presents the bias for the common coefficient case in which $\sigma_\alpha = 0$ for the unmatched and matched comparison groups, respectively. Switching from the variable coefficient case to the common coefficient case has two important effects. First, in the common coefficient case selection into the program depends solely on θ_i and U_{it}. In contrast, in the random coefficient base case, persons with values of θ_i or U_{it} near zero, or even positive, will nonetheless select into training if they have a large enough value of α_i. Figs. 12 and 13 show that in the common coefficient case, the distribution of θ_i for trainees differs much more sharply from that for non-trainees than in the random coefficient base case. A further consequence of eliminating α_i as a determinant of program participation is that Ashenfelter's dip becomes much deeper in the common coefficient case, reflecting the stronger sorting on θ_i and U_i. Figs. 14 and 15 illustrate this difference.

In the random coefficient base case, selection into the program based on α_i acts like randomization for the parameter $\mathrm{E}(\alpha_i \mid D_i = 1)$ because α_i is uncorrelated with all of the components of post-program error. The more D_i is driven by variation in α_i, the more exogenous it is and the smaller the bias. To see this, compare the cross-section estimator in columns (1) and (2). Without the benefit of the pseudo-randomization induced by selection into the program based on α_i, the bias in the common coefficient case, which has the same variances of θ_i and U_{it} as the base case, is much greater. The stronger selection on θ_i and U_i in the common coefficient case and the deeper dip it induces substantially increase the mean bias in all cases except the IV and Heckman (1979) estimators for both the unmatched and matched comparison groups.

The second important effect of switching to the common coefficient model is to dramatically improve the performance of the IV and Heckman (1979) estimators. (This also shows up in their excellent performance in column (4) for the model in which $\mathrm{E}(\alpha_i \mid D_i = 1) = \mathrm{E}(\alpha_i)$ and there is no selection on α_i.) As discussed in Section 7.4, in the common coefficient case, Z_i is a valid instrument (or exclusion restriction) because it is no longer correlated with the error term conditional on D_i. As a result, both the mean bias and the variability in the estimates across samples fall.

Column (3) of Tables 10 and 11 shows the mean bias when $\mathrm{E}(\alpha_i)$ rather than $\mathrm{E}(\alpha_i \mid D_i = 1)$ is the parameter of interest. As indicated by the values in the column headings these parameters differ greatly in our base case model because there is strong selection into the program of persons with high values of α_i. As a result, estimators which estimate $\mathrm{E}(\alpha_i \mid D_i = 1)$ with low bias provide highly biased estimates of $\mathrm{E}(\alpha_i)$. For this parameter, the dependence of D_i on α_i is a source of bias rather than a solution to the bias problem as it is when the parameter of interest is $\mathrm{E}(\alpha_i \mid D_i = 1)$.

Column (4) of Tables 10 and 11 shows the bias for the case where α_i varies across

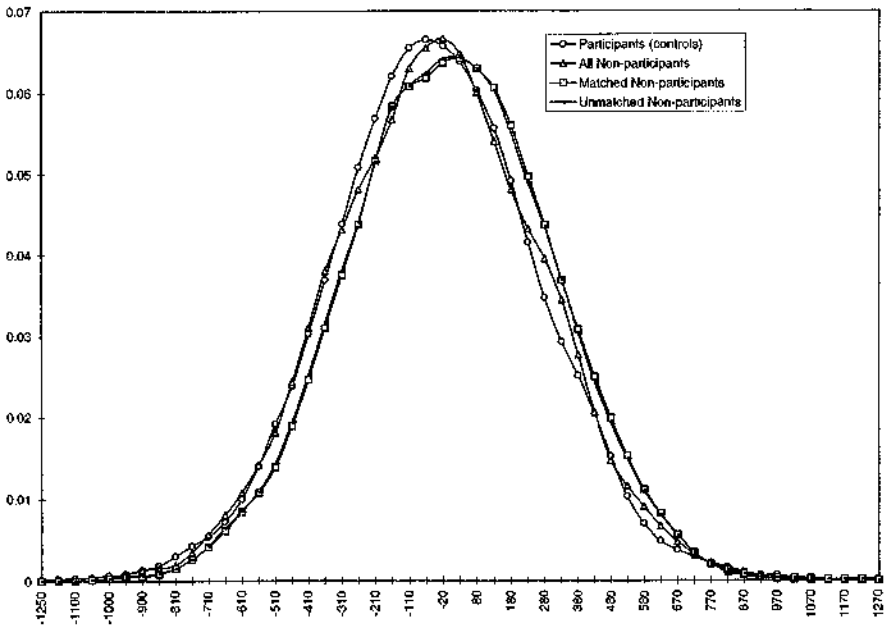

Fig. 12. Distribution of θ in base case with random coefficient.

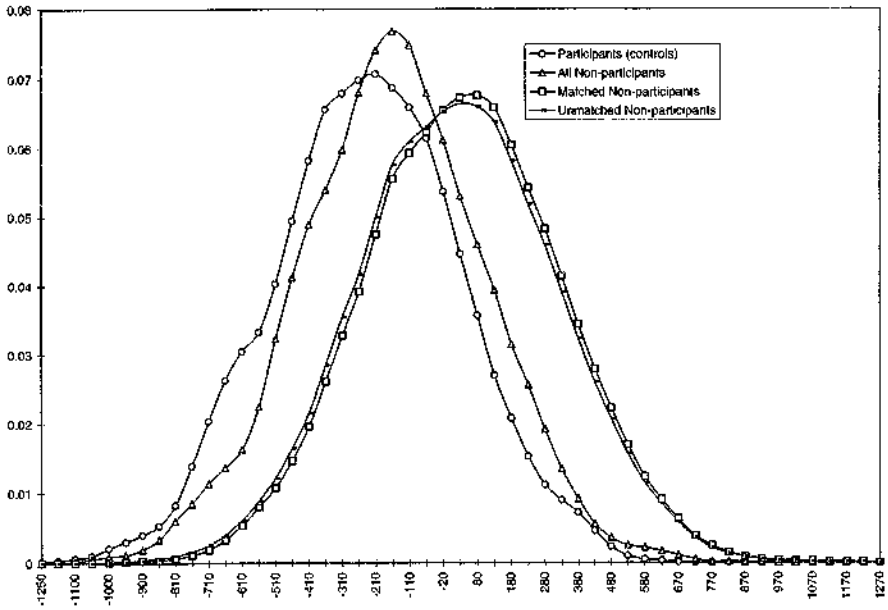

Fig. 13. Distribution of θ in base case with common coefficient.

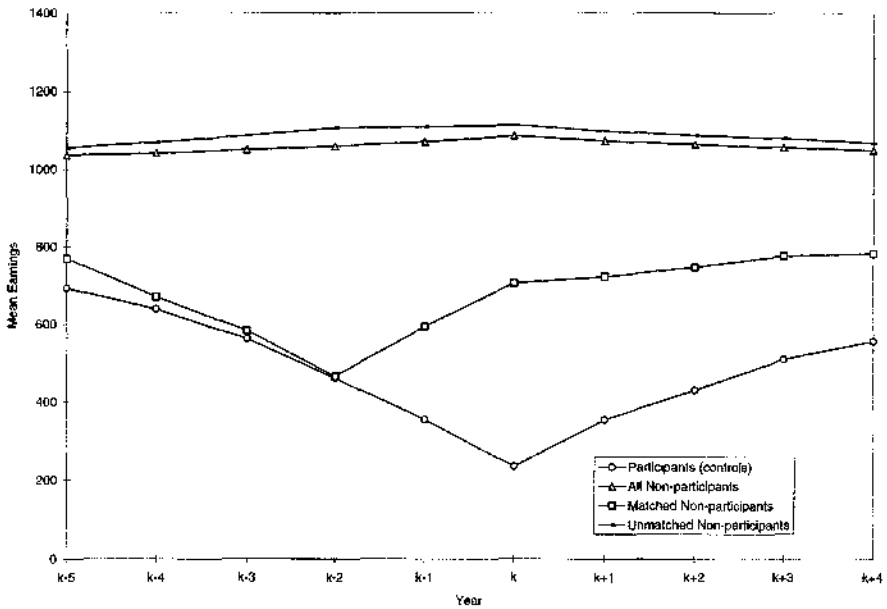

Fig. 14. Mean earnings in base case with common coefficient and matching on Y_{k-2}.

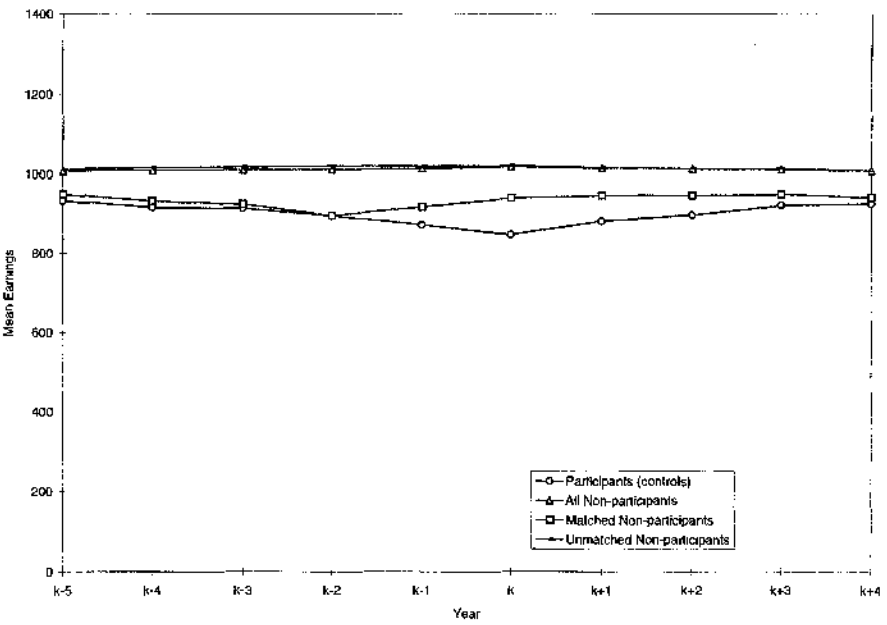

Fig. 15. Mean earnings in base case with random coefficient and matching on Y_{k-2}.

persons but selection decisions are based only on the expected value $E(\alpha_i)$. That is, in this case, potential trainees are assumed not to know the idiosyncratic component of their gain (or loss) from training. The estimated biases for the all of the estimators other than the AR(1) are essentially the same as in the common coefficient case, because in both cases variation in D_i is not driven by variation in α_i. For the AR(1) estimator, the additional error component in the earnings equation adversely affects the performance of the estimator.

The final column of Tables 10 and 11 presents the estimates when the base case is altered by increasing the variance of α_i while holding the variances of θ_i and U_{it} fixed. This essentially randomizes D_i against the error term $\theta_i + U_{it} + D_i(\alpha_i - E(\alpha_i \mid D_i = 1))$. Ceteris paribus, increasing the heterogeneity of the impact of treatment improves the performance of all of the estimators we examine except for the IV and Heckman (1979) estimators so long as $E(\alpha_i \mid D_i = 1)$ is the parameter of interest and agents act on α_i in making program participation decisions.

The evidence presented in Table 10 is based on Monte Carlo using simulated samples of 1000 observations. Though small, evaluators often use samples of this size in practice. In order to study how much of the bias reported in Table 10 results from failure to converge to the true bias values, and in order to gauge the reliability of large sample theory when applied to samples of the sizes used in practice, we present bias estimates from simulated samples of size 2500, 5000 and 10,000 in Table 12. The estimates of bias reported there correspond to those reported in column (1) of Table 10. We find that the estimates in Table 10 provide an accurate gauge of the bias present in all of the non-experimental estimators we examine other than the IV estimator. The IV estimator constitutes an important exception because it converges slowly and is unstable in small samples.

Table 13 shows the effects on the estimated bias for the base case model presented in Table 10 of changing the relative variances of the observed and unobserved variables affecting earnings and participation. The first two columns vary the contributions of U_{it} and θ_i to the outcome equation error variance, holding the overall variance, $\mathrm{Var}(U_{it}) + \mathrm{Var}(\theta_i)$, fixed. Columns (3) and (4) vary the relative contributions of Z_i and V_i in determining D_i. The final column presents the special case where there is no selection bias in post-program outcomes because $\mathrm{Var}(\theta_i) = 0$ and $\rho = 0$. In each case, the exact values for the variances appear in the table notes.

The results reported in the first two columns are pretty much as expected. The bias for the cross-section estimator increases when the contribution of θ_i to the outcome equation increases. The difference-in-differences estimators are designed for the case where θ_i is an important component of the bias. A comparison of columns (1) and (2) reveals that as the variance of θ_i increases, the bias from using this estimator decreases. The AR(1) estimator is designed to exploit the autoregressive properties of the error term in the outcome equation. Therefore it is not surprising that as the variance due to the autoregressive component declines and the variance due to the fixed effect increases, the performance of the AR(1) estimator deteriorates. The performance of the other estimators is not much affected by the relative variances of θ_i and ε_i. This is not surprising because they do not depend on the time series properties of the error terms in the outcome equation.

The second two columns present the bias in the base case model when the relative variances of the observables, Z, and unobservables, V, in the participation equation are changed, keeping the total variance fixed. These changes affect only the IV and Heckman (1979) estimators, which make explicit use of the participation equation. As the variance of Z declines from column (4) to column (3), the correlation of D with Z drops from 0.0677 to 0.0063, and the quality of both the IV and Heckman (1979) estimators declines, as evidenced by the increases in mean bias and in the variance of the bias across the simulated samples. Both estimators rely on an exclusion restriction and on variation in Z_i relative to the outcome equation error term, although they use this information in different ways. As a result, when the exogenous variation in Z_i is small, the performance of these estimators deteriorates.

The case of no selection bias shown in column (5) of Table 13 is an ideal case for all of the estimators other than the Ashenfelter (1979) estimator, which makes use of Y_{ik}. As expected, almost all of the estimators show a very low estimated mean bias. However, it is surprising that the IV estimator does so poorly in this case. This poor performance reflects the intrinsic variability in the IV estimator already noted in our discussion of Table 12.

Tables 14 and 15 indicate the sensitivity of the estimated biases to different matching rules when the matched comparison group samples are used. Table 14 reports estimates for the base case and Table 15 reports estimates for the common coefficient case. The first four columns present biases from matching on earnings at different lags, or on the sum of earnings over the five pre-program periods. The final column reports estimates based on matching on a propensity score obtained by estimating a probit model of participation including Z_i, $Y_{i,k-1}$, $Y_{i,k-2}$ and $Y_{i,k-3}$ as independent variables.

As noted in Section 7.2, using matching to construct a comparison sample alters the properties of the generated samples compared to random samples and thereby affects the properties of many estimators. The Heckman (1979) estimator is especially vulnerable because matching alters the joint distribution of the unobservables in the participation and outcome equations. The IV estimator is also sensitive to departures from random sampling for reasons analyzed in Section 7.7. In both cases, Table 15, as well as the estimates reported in Table 11 for matched samples, demonstrates that these effects are especially pronounced in the case of the common coefficient model where $\sigma_\alpha = 0$, as the variability in the bias from both estimators is substantially higher in the common coefficient case. As we have stressed throughout this section, and as is already evident in column (3) of Tables 10 and 11, increasing the variance of α_i when the parameter of interest is $\mathrm{E}(\alpha_i \mid D_i = 1)$, and persons select into the program on the basis of α_i (and other variables), reduces bias because more of the variation in D_i results from factors that do not contribute to selection bias.

8.4. Specification testing and the fallacy of alignment

The message of Sections 3–7 is that the choice of an estimator to evaluate a program requires making judgments about outcome equations, participation rules and the rela-

Table 14
Bias in non-experimental estimates of the impact of training (matched comparison group samples): parameter of interest: $E(\alpha \mid D = 1) = 615.7$[a]

Estimator[b]	Base case[c] with matching on Y_{k-3} (1)	Base case with matching on Y_{k-2} (2)	Base case with matching on Y_{k-1} (3)	Base case with matching on sum of Y_{k-5} to Y_{k-1} (4)	Base case with propensity score matching (5)
Cross-section					
Mean	−55.8	−42.9	−16.2	−38.4	−30.4
SD	(71.3)	(80.3)	(75.4)	(72.1)	(73.8)
Diff-in-diff (−1,3)					
Mean	2.6	−5.8	−28.1	−12.4	−40.0
SD	(68.7)	(77.5)	(62.9)	(75.3)	(70.8)
Diff-in-diff (−3,3)					
Mean	−64.5	−43.3	−22.3	−57.8	−40.7
SD	(65.9)	(82.5)	(78.3)	(82.7)	(74.6)
Diff-in-diff (−5,3)					
Mean	−52.6	−33.9	−14.6	−53.4	−31.6
SD	(81.5)	(80.3)	(69.4)	(79.0)	(81.9)
AR(1) regression					
Mean	14.7	−6.9	75.6	22.8	45.8
SD	(315.5)	(333.7)	(324.8)	(345.4)	(366.9)
IV estimator					
Mean	346.1	−305.0	399.2	−192.1	1745.8
Median	−242.5	−41.9	−329.4	−49.5	463.0
SD	(3106.5)	(4001.6)	(26174.9)	(5609.5)	(12488.3)
Corr(Z,D)	0.0937	0.0923	0.0926	0.0931	0.0045

Ashenfelter (1979)					
Mean	80.4	81.5	86.1	77.4	66.6
SD	(80.2)	(83.0)	(83.1)	(73.2)	(71.2)
Heckman (1979)					
Mean	123.5	−15382.5	−4625.8	−93.3	2443.1
Median	−77.4	−238.5	−228.8	81.7	718.4
SD	(3508.1)	(155427.6)	(33888.8)	(7386.2)	(11799.4)
Kitchen sink					
Mean	−19.3	−22.8	−12.4	−21.9	−30.4
SD	(70.3)	(80.2)	(77.0)	(70.1)	(72.6)

[a] Estimates are based on 100 simulated samples of 1000 observations each. The "mean" row presents the mean of the estimates from the 100 samples while the "SD" row presents the standard deviation of the estimates from the 100 samples. The "Corr(Z,D)" row for the IV estimates gives the average correlation between the participation indicator, D, and the instrument, Z. Matching consists of nearest neighbor matching with replacement in all cases. In Columns (1), (2) and (3), matching is on earnings in periods $k-3$, $k-2$ and $k-1$, respectively, where period k is the period of participation for those taking training. In Column (4), matching is on the sum of earnings in the five periods prior to period k. In the final column, matching is on a propensity score calculated by estimating a probit model with participation as the dependent variable and Z, Y_{k-3}, Y_{k-2} and Y_{k-1} as independent variables. The average number of unique observations in a matched sample is 92.1 in column (1), 92.1 in column (2), 92.2 in column (3), 91.7 in column (4) and 91.3 in column (5).

[b] The cross-section estimator is the simple difference between participant and non-participant earnings in period $k+4$. The difference-in-differences estimates are based on the periods indicated, so that $(-1,3)$ is the difference between the change in participant earnings from period $k-1$ to period $k+3$ and the change in non-participant earnings over the same interval. The difference-in-differences $(-3,3)$ estimator is symmetric. The AR(1) estimates are based on a regression of Y_{k+4} on Y_{k-3} and D, with the estimate consisting of the coefficient estimate on D divided by $(1-\rho)$, where ρ is estimated by the coefficient on Y_{k+3}. The IV estimates use Z as an instrument for a regression of Y_{k+4} on D. The Ashenfelter (1979) estimator is described in Section 8.3.3. The dependent variable for this estimator is $Y_{k+4}-Y_k$. The Heckman (1979) estimator is a special case of the class of control function estimators presented in Section 7.4.2. The estimates in all five columns are calculated as shown in Section 7.4.2. The dependent variable for the Heckman (1979) estimator is Y_{k+4}. The kitchen sink estimates are based on a regression of Y_{k+4} on Y_{k-1}, Y_{k-2} and Z.

[c] The "base case", has $\theta \sim N(0,300)$, $\varepsilon \sim N(0,280)$, $Z \sim N(0,300)$, $V \sim N(0,200)$, $\rho = 0.78$, and $\alpha = 100 + N(0,300)$. Estimates for the base case without matching appear in Table 10. This case is based on estimates of the size of the permanent and transitory components of earnings from Ashenfelter and Card (1985) and of the variance in the impacts of training from Heckman et al. (1997c). In the base case, the fractions of $\mathrm{Var}(Y_{k+4} \mid D=1)$ accounted for by α and θ are 0.0564 and 0.2670, respectively.

Table 15
Bias in non-experimental estimates of the impact of training (matched comparison group samples); parameter of interest: $E(\alpha \mid D = 1) = 100.0$[a]

Estimator[b]	Base case[c] with common coefficient and matching on Y_{k-3} (1)	Base case with common coefficient and matching on Y_{k-2} (2)	Base case with common coefficient and matching on Y_{k-1} (3)	Base case with common coefficient and matching on sum of Y_{k-5} to Y_{k-1} (4)	Base case with common coefficient and propensity score matching (5)
Cross-section					
Mean	−287.4	−233.0	−165.3	−199.5	−194.6
SD	(72.3)	(70.4)	(80.0)	(73.2)	(103.6)
Diff-in-diff (−1.3)					
Mean	38.3	−36.8	−178.7	−33.0	−206.8
SD	(82.0)	(77.1)	(80.9)	(74.4)	(100.7)
Diff-in-diff (−3.3)					
Mean	−333.6	−243.2	−157.8	−296.5	−214.5
SD	(72.5)	(70.9)	(88.2)	(74.9)	(115.9)
Diff-in-diff (−5.3)					
Mean	−271.1	−202.0	−141.8	−302.6	−177.9
SD	(76.6)	(76.1)	(87.0)	(71.4)	(108.9)
AR(1) regression					
Mean	103.0	−69.1	−150.4	−144.6	37.6
SD	(1124.9)	(479.3)	(1010.3)	(2357.7)	(567.5)
IV estimator					
Mean	−35.8	27.3	89.4	42.4	9387.0
Median	−18.9	18.1	89.1	49.5	550.3
SD	(167.5)	(175.8)	(162.4)	(165.6)	(86337.9)
Corr(Z,D)	0.4781	0.5035	0.5398	0.5029	0.0109

Ashenfelter (1979)					
Mean	223.1	209.3	171.7	215.2	242.7
SD	(85.6)	(85.0)	(77.1)	(86.7)	(109.6)
Heckman (1979)					
Mean	− 32.3	24.1	81.2	42.5	9499.4
Median	− 23.4	13.9	80.1	52.1	559.7
SD	(167.2)	(177.2)	(171.0)	(167.0)	(87144.0)
Kitchen sink					
Mean	− 161.5	− 194.7	− 201.8	− 162.0	− 187.5
SD	(85.6)	(91.5)	(99.3)	(101.1)	(98.6)

[a] Estimates are based on 100 simulated samples of 1000 observations each. The "mean" row presents the mean of the estimates from the 100 samples while the "SD" row presents the standard deviation of the estimates from the 100 samples. The "Corr(Z,D)" row for the IV estimates gives the average correlation between the participation indicator, D, and the instrument, Z. Matching consists of nearest neighbor matching with replacement in all cases. In Columns (1), (2) and (3), matching is on earnings in periods $k-3$, $k-2$ and $k-1$, respectively, where period k is the period of participation for those taking training. In Column (4), matching is on the sum of earnings in the five periods prior to period k. In the final column, matching is on a propensity score calculated by estimating a probit model with participation as the dependent variable and Z, Y_{k-3}, Y_{k-2} and Y_{k-1} as independent variables. The average number of unique observations in a matched sample is 84.0 in column (1), 79.8 in column (2), 73.0 in column (3), 78.7 in column (4) and 56.3 in column (5).

[b] The cross-section estimator is the simple difference between participant and non-participant earnings in period $k+4$. The difference-in-differences estimates are based on the periods indicated, so that (−1,3) is the difference between the change in participant earnings from period $k-1$ to period $k+3$ and the change in non-participant earnings over the same interval. The difference-in-differences (−3,3) estimator is symmetric. The AR(1) estimates are based on a regression of Y_{k-4} on Y_{k-3} and D, with the estimate consisting of the coefficient estimate on D divided by $(1-\rho)$, where ρ is estimated by the coefficient on Y_{k-3}. The IV estimates use Z as an instrument for a regression of Y_{k+4} on D. The Ashenfelter (1979) estimator is described in Section 8.3.3. The dependent variable for this estimator is $Y_{k+4}-Y_k$. The Heckman (1979) estimator is a special case of the class of control function estimators presented in Section 7.4.2. The estimates in all five columns consist of the coefficient on D when the estimated control functions are included. The dependent variable for the Heckman (1979) estimator is Y_{k+4}. The kitchen sink estimates are based on a regression of Y_{k+4} on Y_{k-1}, Y_{k-2} and Z.

[c] The "base case", has $\theta \sim N(0,300)$, $\varepsilon \sim N(0.450)$, $Z \sim N(0,300)$, $V \sim N(0,200)$, $\rho = 0.78$, and $\alpha = 100$. Estimates for the base case without matching appear in Table 10. This case is based on estimates of the size of the permanent and transitory components of earnings from Ashenfelter and Card (1985). In the base case, with common coefficient, the fractions of $\mathrm{Var}(Y_{k-4} \mid D=1)$ accounted for by α and θ are 0.0000 and 0.3132, respectively.

tionship between the two. All estimators, including social experiments, are based on identifying assumptions which are often difficult if not impossible to test on the available data. For example, the validity of social experiments depends on assumption (5.A.1) or assumptions (5.A.2a) and (5.A.2b), which state that randomization does not disrupt the program being evaluated. Testing for disruption effects turns out to be a difficult task (see Heckman et al., 1996a). Testing whether a variable is a valid instrument is also difficult unless one has access to the true parameter via some other identifying assumption, such as another instrument, a valid social experiment or one of the other identifying restrictions discussed above or in Heckman and Robb (1985a, 1986a). The inability to test maintained identifying assumptions on the available data is a source of frustration to many.

One widely used practice in the evaluation literature apparently evades this problem by testing evaluation models on pre-program data and then using the models that pass the tests to evaluate the program. Papers by Ashenfelter (1978), Ashenfelter and Card (1985) and Heckman and Hotz (1989) exemplify this approach. The idea underlying this approach is that if a selection estimator correctly adjusts for differences in pre-program earnings levels (or some other outcome measure) between future participants and non-participants, it should also adjust correctly for post-program differences and therefore be a valid estimator for evaluating the program. This method could also be applied to the matching estimators defined in Section 7.4.1. According to this line of reasoning, a good match on pre-program outcome levels should produce a valid estimator for post-program levels.

The basic idea underlying this method is captured by the following testing framework. Write $A(Y_{1t'},X_{t'})$ for the adjusted pre-program earnings of program participants and $A(Y_{0t'},X_{t'})$ for the adjusted pre-program earnings of non-participants, where $t' < k$. Then, for a common $X_{t'}$, test the hypothesis

$$A(Y_{1t'},X_{t'}) = A(Y_{0t'},X_{t'}). \tag{8.10}$$

Most commonly such tests are based on the model of (3.10). In that context, the test for a valid comparison group is a test of the hypothesis H_0: $\alpha = 0$ in the equation

$$Y_{t'} = X_{t'}\beta + D\alpha + U_{t'}, \qquad t' < k,$$

estimated using pre-program data on participants and comparison group members. Here $D = 1$ denotes that a person will be a participant in period k. If H_0 is not rejected, the comparison group is deemed to be adequate.

This logic seems compelling, but is potentially misleading. The success of testing strategies based on the alignment of pre-program earnings depends on the serial correlation properties of the error term in the earnings equation. Suppose, for example, that program participants and non-participants have identical pre-program earnings histories but that participants experience a permanent loss in earnings at the time of enrollment in period k. In this case, finding that a particular estimator or comparison group correctly aligns earnings in periods prior to k tells little about the validity of a post-

program comparison. Even if the program had a strong positive impact on participant earnings compared to what they would have earned without the program, post-program comparisons between participants and non-participants based on estimators or comparison groups which correctly aligned pre-program earnings might still yield a negative impact estimate for the program because of the large negative shock experienced by participants.

Using tests based on the alignment of pre-program earnings or outcome levels to evaluate the validity of an estimator or comparison group or both is the alignment fallacy. The widely used Heckman and Hotz (1989) tests of the validity of non-experimental selection estimators using pre-program earnings are based on the alignment fallacy. Its practical importance can be illustrated by re-examining an old controversy in the evaluation literature. In the early 1980s, two major consulting firms – Westat, Inc. and SRI International – used matching to construct comparison groups to evaluate the US CETA training program. Both firms had access to the same large datasets and both hired expert statisticians who advocated matching as an evaluation estimator. They both chose their comparison groups to align the earnings of participants and comparison group members in the pre-program period.

As shown in Fig. 3, Ashenfelter's dip characterized the earnings of participants in the CETA program. SRI chose to match on earnings two periods prior to the enrollment period. It picked as comparison group members persons whose earnings were very similar to participants in period $k-2$. Westat aligned using earnings in period $k-1$. Using a simple matching estimator for post-program earnings, SRI reported a negative impact of CETA on participant earnings that was substantially lower than the impact reported by Westat. Figs. 15 and 16 demonstrate how this would happen. Those figures are based on our adaptation of the empirical model of Ashenfelter and Card (1985) used to generate the simulations in Section 8.3. That model is rich enough to generate Ashenfelter's dip. Figs. 15 and 16 show the earnings of participants, matched non-participants, unmatched non-participants and all non-participants for comparison groups based on matching in periods $k-2$ and $k-1$, respectively.

Comparing Figs. 15 and 16, when we match so that future participant and non-participant earnings are the same in period $k-2$, mean reversion causes the earnings after period k of persons aligned in $k-2$ to be higher than those of persons aligned based on earnings in period $k-1$.[88] This implies that the matching estimator used by SRI should produce a lower estimate of program impact than the matching estimator used by Westat, which is exactly what was found. Neither matching estimator may be correct, but the ordering of the estimates obtained from them is predicted from our knowledge of the earnings dynamics of program participants.

Alignment on pre-program earnings is not guaranteed to produce valid estimators of the impact of a program using post-program earnings. It is thus interesting, but not by any

[88] There were other matching variables used by both groups but the use of earnings at different lags to form matched samples plays the main role in explaining the discrepancy between the two studies.

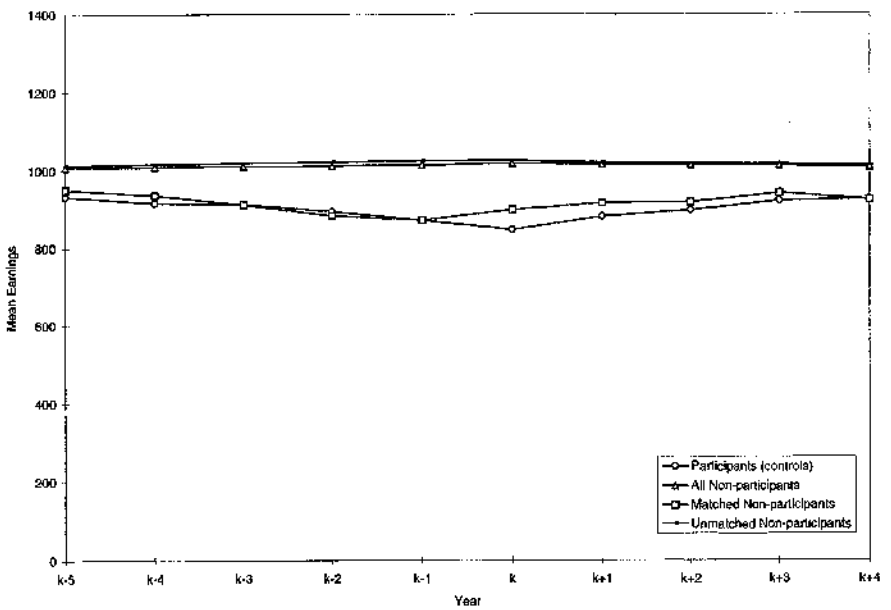

Fig. 16. Mean earnings in base case with random coefficient and matching on Y_{k-1}.

means conclusive, that specification tests based on alignment of pre-program earnings developed by Heckman and Hotz (1989) have been found by them and by others such as Friedlander and Robbins (1995) to eliminate from consideration the most biased estimators of training impact. Even in these studies, many estimators that survive the tests still exhibit substantial bias.

8.4.1. Testing identifying assumptions

As noted by Heckman and Robb (1985a, 1986a), most of the conventional econometric estimators make strong overidentifying restrictions which can be tested. The fixed effects and inverse Mills' ratio estimators are examples of evaluation models with strong overidentifying assumptions.[89] Heckman et al. (1997a) present tests of over-identifying assumptions for matching estimators for non-experimental data.

Nonetheless, Heckman and Robb (1985a, 1986a) also note that all econometric evaluation models can be weakened to a just-identified form, and they present many examples of how this can be done. Just-identified models offer one interpretation of the available data but other just-identified models are equally good descriptions of the same data. The only

[89] Tests of the fixed effect model for panels of length greater than $T = 2$ are presented in Chamberlain (1984), Hsiao (1986) and Baltagi (1995). Tests for the normal selection model based on the properties of censored normal residuals are discussed in Amemiya (1985) among other sources. See also Bera et al. (1984).

way to test the validity of just-identified models is to get better data to eliminate the effects of unobservables on selection.

9. Indirect effects, displacement and general equilibrium treatment effects

Except for our discussion of general equilibrium effects in Section 3.4, throughout this chapter we have followed most of the evaluation literature and used microeconomic partial equilibrium analysis as a framework for interpreting the estimates obtained from evaluation studies. As stated in Section 3, the key identifying assumption in this approach is that the no-treatment outcomes within a given policy regime closely approximate the outcomes in a no-program regime. In the language of Lewis (1963), this assumption allows analysts to ignore indirect effects. In the context of evaluating large scale employment and training programs at a national level, it is natural to ask whether this assumption is valid and the consequences for an evaluation if it is not. To answer these questions in a convincing fashion requires constructing a model of the labor market, a task that is rarely performed in conventional evaluation studies.

In this section, we summarize a line of previous research that attempts to unite the "treatment effect" literature with the general equilibrium policy evaluation literature. Calls for doing so originate in the work of Lewis (1963) and have also been made by Hamermesh (1971, 1993), Johnson and Layard (1986) and others. Within the framework of a Mortensen–Pissarides model, Davidson and Woodbury (1993, 1995) present a promising attempt to analyze the indirect effects of an unemployment bonus program. They assume that prices and wages are fixed and consider the effects of the bonus program on the search behavior of participants and non-participants. In a model with flexible skill prices, Heckman et al. (1998d) consider the effects of changes in tuition on schooling and earnings, accounting for general equilibrium effects on participants and non-participants. We consider both models in this section after briefly surveying the traditional approach to accounting for indirect effects.

Newly trained workers may displace previously trained workers if wages are inflexible, as they are in many European countries. For some training programs in Europe, substantial displacement effects have been estimated (OECD, 1993; Calmfors, 1994). If wages are flexible, the arrival of newly trained workers to the market tends to lower the wages of previously trained workers but does not displace any worker. In the framework of Section 3, even if the effect of treatment on the treated is positive, non-participants may be worse off as a result of the program compared to what they would have experienced in the no-program state. Non-participants who are good substitutes for the new trainees will be especially affected. Complementary factors will benefit. These spillover effects can have important consequences for the interpretation of traditional evaluation parameters. The benchmark "no-treatment" state is actually affected by the program.[90]

[90] Thus assumption (3.15) may be violated and instead $E(Y_0 \mid D = 0, \tilde{\varphi}) < E(Y_0 \mid D = 0, \varphi = 0)$.

To demonstrate these issues in a dramatic way, consider the effect of a wage subsidy for employment in a labor market for low-skill workers. Assume that firms act to minimize their costs of employment. Wage subsidies operate by taking non-employed persons and subsidizing their employment at firms.

As indicated in Table 2, many active labor market policies have a substantial wage-subsidy component. Suppose that the reason for non-employment of low-skill workers is that minimum wages are set too high. This case is a traditional justification for wage subsidies (see, e.g., Johnson, 1979; Johnson and Layard, 1986). If the number of subsidized workers is less than the number of workers employed at the minimum wage, a wage subsidy financed from lump sum taxes has no effect on total employment in the low wage sector because the price of labor for the marginal worker hired by firms is the minimum wage which is the same before and after the subsidy program is put in place. The marginal worker is unsubsidized both before and after the subsidy program is put in place.

The effects of the program are dramatic on the individuals who participate in it. Persons previously non-employed become employed as firms seek workers who carry a wage subsidy. Many previously-employed workers become non-employed as their employment is not subsidized. There are no effects of the wage subsidy program on GDP unless the taxes raised to finance the program have real effects on output. Yet there is substantial redistribution of employment. Focusing solely on the effects of the program on subsidized workers greatly overstates its impact on the economy at large.

In order to estimate the impact of the program on the overall economy, $A(\tilde{\varphi})$ in the notation of Section 3, it is necessary to look at outcomes for both participants and non-participants. Only if the benefits accruing to previously-non-employed participants are adopted as the appropriate criterion would the effect of treatment on the treated be a parameter of interest in this situation. Information on both direct participants and affected non-participants is required to estimate the net gain in earnings and employment resulting from the program.

In the case of a wage subsidy, comparing the earnings and employment of subsidized participants during their subsidized period to their earnings and employment in the pre-subsidized period can be a very misleading estimator of the total impact of the program. So is a cross-section comparison of participants and non-participants. In the example of a subsidy in the presence of a minimum wage, the before–after estimate of the gain exceeds the cross-section estimate unless the subsidy is extended to a group of non-employed workers as large as the number employed at the minimum wage. For subsidy coverage levels below this amount, some proportion of the unsubsidized employment is paid the minimum wage. Under these circumstances, commonly-used evaluation estimators produce seriously misleading estimates of program impacts.

The following example clarifies and extends these points to examine the effect of displacement on the trilogy of estimators discussed in Section 4. Let N be the number of participants in the low-wage labor market. Let N_E be the number of persons employed at the minimum wage M and let N_S be the number of persons subsidized. Subsidized persons receive the minimum wage. Subsidization operates solely on persons who would other-

wise have been non-employed and had no earnings. Assume $N_E > N_S$. Therefore, the subsidy has no effect on total employment in the market, because the marginal cost of labor to a firm is still the minimum wage. Workers with the subsidy are worth more to the firm by the amount of the subsidy S. Firms would be willing to pay up to $S + M$ per subsidized worker to attract them.

The estimated wage gain using a before–after comparison for subsidized participants is

$$\text{Before-after}: \qquad S + M - 0 = S + M,$$

because all subsidized persons earn a zero wage prior to the subsidy. The estimated wage gain using cross-section comparisons of program participants and non-participants is

$$\text{Cross-section}: \qquad (S + M) - M\frac{(N_E - N_S)}{(N - N_S)} = S + M\left(\frac{N - N_E}{N - N_S}\right) < S + M,$$

where $(S+M)$ is the average participant's wage and $M(N - N_E)/(N - N_S)$ is the average non-participant's wage. Since $N_E > N_S$, the before–after estimator is larger than the cross-section estimator. The widely used difference-in-differences estimator compares the before–after outcome measure for participants to the before–after outcome measure for non-participants.

Difference-in-differences :

$$(S + M - 0) - M\left(\frac{N_E - N_S}{N - N_S} - \frac{N_E}{N - N_S}\right) = S + M\left(\frac{N}{N - N_S}\right) > S + M.$$

The gain estimated from the difference-in-differences estimator exceeds the gain estimated from the before–after estimator which in turn exceeds the gain estimated from the cross-section estimator. The "no-treatment" benchmark in the difference-in-differences model is contaminated by treatment.

The estimate of employment creation obtained from the three estimators is obtained by setting $M = 1$ and $S = 0$ in the previous expressions. This converts those expressions into estimates of employment gains for the different groups used in their definition.

None of these estimators produces a correct assessment of wage or employment gain for the economy at large. Focusing only on direct participants causes analysts to lose sight of overall program impacts. Only an aggregate analysis of the economy as a whole, or random samples of the entire economy, would produce the correct assessment that no wage increase or job creation is produced by the program. The problem of indirect effects poses a major challenge to conventional micro methods used in evaluation research that focus on direct impacts instead of total impacts, and demonstrates the need for program evaluations to utilize market-wide data and general equilibrium methods.

9.1. Review of the traditional approaches to displacement and substitution

Calmfors (1994) presents a comprehensive review of the issues that arise in evaluating

active labor market programs in the context of a modern economy and an exhaustive list of references on theoretical and empirical work on this topic. He distinguishes a number of indirect effects including *displacement effects* (jobs created by one program are at the expense of other jobs), *deadweight effects* (subsidizing hiring that would have occurred in the absence of the program), *substitution effects* (jobs created for a certain category of workers replace jobs for other categories because relative wage costs have changed) and *tax effects* (the effects of taxation required to finance the programs on the behavior of everyone in society). A central conclusion of this literature is that the estimates of program impact from the microeconomic treatment effect literature provide incomplete information about the full impacts of active labor market programs. The effect of a program on participants may be a poor approximation to the total effect of the program, as our simple example has shown.

Forslund and Krueger (1997) illustrate both the traditional approach to estimating displacement and the problems with it. The standard reduced form approach pursued by Johnson and Tomola (1977), Gramlich and Ysander (1981) and others regresses employment in non-subsidized jobs in a geographical area on the number of subsidized jobs lagged one period and other control variables. Full displacement is said to occur if the estimated coefficient on lagged subsidized employment is minus one. For each subsidized job there is one fewer unsubsidized job. For Swedish construction workers, Forslund and Krueger estimate a coefficient of -0.69 so that for each public relief worker hired, there are 0.69 fewer private construction workers hired. For other groups, their estimates of displacement are unstable and they report only a broad range of values.

Forslund and Krueger discuss the problem of reverse causation. A negative shock to the economy may stimulate the use of relief workers. The estimated displacement effect may be a consequence of the feedback between macro shocks and the application of a public hiring policy. Although they present various ad hoc methods based on vector autoregressions to circumvent this problem, they sound a cautionary note about all of the reduced form methods used to estimate displacement and the evidence presented in the entire literature based on them.

9.2. *General equilibrium approaches*

A more clearly interpretable approach to the problem of measuring indirect effects of programs is to construct equilibrium models of the labor market in which both direct and indirect effects are modeled. One recent example is Davidson and Woodbury (1993). They consider these issues in the context of evaluating a bonus scheme to encourage unemployed workers to find jobs more quickly using a Mortensen–Pissarides search model in which prices are fixed. A second recent example is the analysis of Heckman et al. (1998e). They consider the evaluation of tuition subsidy programs in a general equilibrium model of human capital accumulation with both schooling and on-the-job training and with heterogeneous skills in which prices are flexible. The first is a model of displacement with fixed prices; the second is a model of substitution.

Both studies demonstrate the problems with, and possibilities for, general equilibrium analysis of the impacts of active labor market programs. They both find important indirect effects of the programs they evaluate. At the same time, both studies demonstrate that the task of finding credible parameters for general equilibrium models is a challenging one. We first consider the analysis of Davidson and Woodbury.

9.2.1. Davidson and Woodbury

The reemployment bonus scheme analyzed by Davidson and Woodbury (1993) accelerates the rate at which unemployed persons offered the bonus find jobs. The bonus is paid to currently unemployed eligible persons with spells below a threshold level who find jobs within a specified time frame. By stimulating aggregate search activity, the bonus may also have macro effects on output and on the search behavior of unsubsidized participants. The higher taxes raised to finance the program may reduce aggregate search activity by the unsubsidized as their return to market activity declines. The higher level of search by the subsidized may discourage search by their unsubsidized competitors in the labor market.

Davidson and Woodbury (1993) consider four classes of workers: (a) unemployment insurance (UI) recipients who are eligible for the bonus if they get hired; (b) UI recipients who are ineligible for a bonus because of the length of their current unemployment spell (the bonus is only paid to persons with an unemployment spell below a certain length); (c) UI recipients who have exhausted their benefits; and (d) jobless workers who were never eligible to receive UI benefits and cannot receive a bonus. They develop an equilibrium model of search assuming that workers are income maximizing and the bonus is offered in the steady state.

Workers eligible for a bonus have an incentive to accelerate their search. Those ineligible for a bonus in the current spell experience two offsetting effects: (a) the competition for jobs increases, making search less profitable and (b) the benefits of being unemployed rise in the next spell because of the bonus. The second effect promotes search because of the eligibility for the program conferred on persons when they eventually secure a job and are at risk for future unemployment. In their simulations these effects cancel out, leaving the search activity of this group unaffected. However, because of enhanced search by those with the subsidy, the rate of job acquisition declines for those currently ineligible for the bonus.

For those who are permanently ineligible, only the first effect operates; as a result they reduce their search activity. This generates displacement. During recessions, the existence of a bonus leads to displacement of non-bonus workers (those permanently ineligible and those whose benefits are exhausted or whose eligibility has expired). Permanently ineligible workers always experience displacement. Davidson and Woodbury estimate that 30–60% of the gross employment effect of the bonus program is offset by displacement of UI-ineligible workers. Microeconomic treatment analyses of program participant employment experiences provide a substantially misleading picture of the effect of the program on society at large. We next turn to a general equilibrium model of an economy with wage flexibility and indirect effects.

9.2.2. Heckman, Lochner and Taber

The typical microeconomic evaluation of tuition policy estimates the response of college enrollment to tuition variation using geographically dispersed cross-sections of individuals facing different tuition rates. These estimates are then used to determine how subsidies to tuition will raise college enrollment. The impact of tuition policies on earnings are evaluated using a schooling-earnings relationship fit on pre-intervention data and do not account for the enrollment effects of the taxes raised to finance the tuition subsidy. Kane (1994) and Cameron and Heckman (1998) exemplify this approach.

The danger in this widely used practice is that what is true for policies affecting a small number of individuals, as studied by social experiments or as studied in the microeconomic "treatment effect" literature, need not be true for policies that affect the economy at large. A national tuition-reduction policy may stimulate substantial college enrollment and will also likely reduce skill prices. However, agents who account for these changes will not enroll in school at the levels calculated from conventional procedures which ignore the impact of the induced enrollment on skill prices. As a result, standard policy evaluation practices are likely to be misleading about the effects of tuition policy on schooling attainment and wage inequality. The empirical question is: how misleading? Heckman et al. (1998e) show that conventional practices in the educational evaluation literature lead to estimates of enrollment responses that are ten times larger than the long-run general equilibrium effects. They improve on current practice in the "treatment effects" literature by considering both the gross benefits of the program and the tax costs of financing the treatment as borne by different groups.

Evaluating the general equilibrium effects of a national tuition policy requires more information than the tuition-enrollment parameter that is the centerpiece of partial equilibrium policy analysis. Policy proposals of all sorts typically extrapolate well outside the range of known experience and ignore the effects of induced changes in skill quantities on skill prices. To improve on current practice, Heckman et al. (1998e) develop an empirically justified rational expectations perfect foresight overlapping-generations general equilibrium framework for the pricing of heterogeneous skills. It is based on an empirically grounded theory of the supply of schooling and post-school human capital, where different schooling levels represent different skills. Individuals differ in learning ability and in initial endowments of human capital. Household saving behavior generates the aggregate capital stock, and output is produced by combining the stocks of different human capitals with physical capital. Factor markets are competitive and there is price flexibility. The framework explains the pattern of rising wage inequality experienced in the United States in the past 30 years. They apply their framework to evaluate tuition policies that attempt to increase college enrollment.

For two reasons, the "treatment effect" framework that ignores the general equilibrium effects of tuition policy is inadequate. First, the parameters of interest depend on who in the economy is "treated" and who is not. Second, these parameters do not measure the full impact of the program. For example, increasing tuition subsidies may increase the earnings of uneducated individuals who do not take advantage of the subsidy. To pay for the

subsidy, the highly educated would be taxed and this may affect their investment behavior. In addition, more competitors for educated workers enter the market as a result of the policy, and their earnings are depressed. Conventional methods ignore the effect of the policy on non-participants operating through changes in equilibrium skill prices as well as Calmfors' tax effect. In order to account for these effects, it is necessary to conduct a general equilibrium analysis.

The analysis of Heckman et al. (1998e) has major implications for the widely used difference-in-differences estimator. If the tuition subsidy changes the aggregate skill prices, the decisions of non-participants will be affected. The "no-treatment" benchmark group is affected by the policy and the difference-in-differences estimator does not identify the effect of the policy for anyone compared to a no-treatment state.[91]

Using their model, Heckman et al. (1998e) simulate the effects on enrollment in college and wage inequality of a revenue-neutral $500 increase in college tuition subsidy on top of existing programs that is financed by a proportional tax. They start from a baseline economy that describes the US in the mid 1980s and that produces wage growth profiles and schooling enrollment and capital stock data that match micro and macro evidence. The partial equilibrium increase in college attendance is 5.3% in the new steady state. This analysis holds skill prices, and therefore college and high school wage rates, fixed – a typical assumption in microeconomic "treatment effect" analyses.

When the policy is evaluated in a general equilibrium setting, the estimated effect falls to 0.46%. Because the college-high school wage ratio falls as more individuals attend college, the returns to college are less than when the wage ratio is held fixed. Rational agents understand this effect of the tuition policy on skill prices and adjust their college-going behavior accordingly. Policy analysis of the type offered in the "treatment effect" literature ignores the responses of rational agents to the policies being evaluated. There is substantial attenuation of the effects of tuition policy on capital and on the stocks of the different skills in their model compared to a partial equilibrium treatment effect model. They demonstrate that their results are robust to a variety of specifications of the economic model.

They also analyze short-run effects. When they simulate the model with rational expectations, the short-run college enrollment effects are also very small, as agents anticipate the effects of the policy on skill prices and calculate that there is little gain from attending college at higher rates. Under myopic expectations, the short-run enrollment effects are much closer to the estimated partial equilibrium effects. With learning on the part of agents, but not perfect foresight, there is still a substantial gap between partial equilibrium and general equilibrium estimates.

Heckman et al. (1998e) also consider the impact of a policy change on discounted earnings and utility and decompose the total effects into benefits and costs, including tax costs for each group, thus isolating Calmfors' tax effect. Table 16 compares outcomes

[91] This problem of spillover effects was first studied by Lewis (1963) who pointed out its implications for estimating the union–non-union wage differential from cross-section and repeated cross-section comparisons.

across two steady states: (a) the benchmark steady state and (b) the steady state associated with the new tuition policy. Given that the estimated schooling response to a $500 subsidy is small, they instead use a $5000 subsidy for the purpose of exploring general equilibrium effects on earnings. (Current college tuition subsidy levels are this high or higher at many colleges in the US.) The row "High School–High School" reports the change in a variety of outcome measures for those persons who would be in high school under either the benchmark or new policy regime; the "High School–College" row reports the change in the same measures for high school students in the benchmark state who are induced to attend college by the new policy; the "College–High School" outcomes refer to those persons in college in the benchmark economy who only attend high school after the policy; and so forth.

By the measure of the present value of earnings, some of those induced to change are worse off. Contrary to the monotonicity assumption built into the LATE parameter discussed in Section 7, and defined in this context as the effect of the tuition subsidy on the earnings of those induced by it to go to college, they find that the tuition policy produces a two-way flow. Some people who would have attended college in the benchmark regime no longer do so. The rest of society also is affected by the policy – again, contrary to the implicit assumption built into LATE that only those who change status are affected by the policy. People who would have gone to college without the policy and continue to do so after the policy are financially worse off for two reasons: (a) the price of their skill is depressed and (b) they must pay higher taxes to finance the policy. However, they now receive a tuition subsidy and for this reason, on net, they are better off both financially and in terms of utility. Those who abstain from attending college in both steady

Table 16
Simulated effects of $5000 tuition subsidy on different groups; steady state changes in present value of lifetime wealth (thousands of 1995 US dollars)[a]

Group (proportion)[b]	After-tax earnings using base tax[c] (1)	After-tax earnings[c] (2)	After-tax earnings net of tuition[c] (3)	Utility[c] (4)
High School–High School (0.528)	9.512	−0.024	−0.024	−0.024
High School–College (0.025)	−4.231	−13.446	1.529	1.411
College–High School (0.003)	−46.711	−57.139	−53.019	−0.879
College–College (0.444)	−7.654	−18.204	0.42	0.42

[a] Source: Heckman et al. (1998e, Table 1).

[b] The groups correspond to each possible counterfactual. For example, the "High School–High School" group consists of individuals who would not attend college in either steady state, and the "High School–College" group would not attend college in the first steady state, but would in the second, etc.

[c] Column (1) reports the after-tax present value of earnings in thousands of 1995 US dollars discounted using the after-tax interest rate where the tax rate used for the second steady state is the base tax rate. Column (2) adds the effect of taxes, column (3) adds the effect of tuition subsidies and column (4) includes the non-pecuniary costs of college in dollar terms.

states are better off in the second. They pay higher taxes, but their skill becomes more scarce and their wages rise. Those induced to attend college by the policy are better off in terms of utility but are not necessarily better off in terms of income. Note that neither category of non-changers is a natural benchmark for a difference-in-differences estimator. The movement in their wages before and after the policy is due to the policy and cannot be attributed to a benchmark "trend" that is independent of the policy.

Table 17 presents the impact of the $5000 tuition policy on the log earnings of individuals with 10 years of work experience for different definitions of treatment effects. The partial equilibrium version given in the first column holds skill prices constant at initial steady state values. The general equilibrium version given in the second column allows prices to adjust when college enrollment varies. Consider four parameters initially defined in a partial equilibrium context. The *average treatment effect* is defined for a randomly selected person in the population in the benchmark economy and asks how that person would gain in wages by moving from high school to college. The parameter *treatment on the treated* is defined as the average gain over their non-college alternative of those who

Table 17
Treatment effect parameters under partial equilibrium and general equilibrium; difference in log earnings for college graduates versus high school graduates at 10 years of work experience[a]

Parameter	Prices fixed[b] (1)	Prices vary[c] (2)	Fraction of sample[d] (%) (3)
Average treatment effect (ATE)	0.281	1.801	100
Treatment on treated (TT)	0.294	3.364	44.7
Treatment on untreated (TOU)	0.270	−1.225	55.3
Marginal treatment effect (MTE)	0.259	0.259	–
LATE[e] 5000 subsidy			
Partial equilibrium	0.255	–	23.6
GE (HS to college) (LATE)	0.253	0.227	2.48
GE (college to HS) (LATER)	0.393	0.365	0.34
GE Net (TLATE)	–	0.244	2.82
LATE[e] 500 subsidy			
Partial equilibrium	0.254	–	2.37
GE (HS to college) (LATE)	0.250	0.247	0.24
GE (college to HS) (LATER)	0.393	0.390	0.03
GE Net (TLATE)	–	0.264	0.27

[a] Source: Heckman et al. (1998e).

[b] In column (1), prices are held constant at their initial steady state levels when wage differences are calculated.

[c] In column (2), we allow prices to adjust in response to the change in schooling proportions when calculating wage differences.

[d] For each row, column (3) presents the fraction of the sample over which the parameter is defined.

[e] The LATE group gives the effect on earnings for persons who would be induced to attend college by a tuition change. In the case of GE, LATE measures the effect on individuals induced to attend college when skill prices adjust in response to quantity movements among skill groups. The partial equilibrium LATE measures the effect of the policy on those induced to attend college when skill prices are held at the benchmark level.

attend college in the benchmark state. The parameter *treatment on the untreated* is defined as the average gain over their college wage received by individuals who did not attend college. The *marginal treatment effect* is defined for individuals who are indifferent between going to college or not. This parameter is a limit version of the LATE parameter under conventional assumptions made in discrete choice theory (Heckman, 1997; Heckman and Vytlacil, 1999a,b). Column 2 presents the general equilibrium version of *treatment on the treated*. It compares the earnings of college graduates in the benchmark economy with what they would earn if no one went to college.[92] The treatment on the untreated is defined analogously by comparing what high school graduates in the benchmark economy would earn if everyone in the population were forced to go to college. The *average treatment effect* compares the average earnings in a world in which everyone attends college versus the earnings in a world in which nobody attends college. Such dramatic policy shifts produce large estimated effects. In contrast, the general equilibrium marginal treatment effect parameter considers the gain to attending college for people on the margin of indifference between attending college and only attending high school. In this case, as long as the mass of people in the indifference set is negligible, partial and general equilibrium parameters are the same.

The final set of parameters Heckman et al. (1998e) consider are versions of the LATE parameter. This parameter depends on the particular intervention being studied and its magnitude. The partial equilibrium version of LATE is defined on the outcomes of individuals induced to attend college, assuming that skill prices do not change. The general equilibrium version is defined for the individuals induced to attend college when prices adjust in response to the policy. The two LATE parameters are quite close to each other and are also close to the marginal treatment effect.[93] General equilibrium effects change the group over which the parameter is defined compared to the partial equilibrium case. For the $5000 subsidy, there are substantial price effects and the partial equilibrium parameter differs substantially from the general equilibrium parameter.

Heckman et al. (1998e) also present partial and general equilibrium estimates for two extensions of the LATE concept: LATER (the effect of the policy on those induced to attend only high school rather than going to college) – Reverse LATE – and TLATE (the effect of the policy on all of those induced to change whichever direction they flow). LATER is larger than LATE, indicating that those induced to drop out of college have larger gains from dropping out than those induced to enter college have from entering. TLATE is a weighted average of LATE and LATER with weights given by the relative proportion of people who switch in each direction.

The general equilibrium impacts of tuition on college enrollment are an order of

[92] In the empirical general equilibrium model of Heckman et al. (1998d), the Inada conditions for college and high school are not satisfied in the aggregate production function and the marginal product of each skill group when none of it is utilized is a bounded number. If the Inada conditions were satisfied, this counterfactual and the counterfactual treatment on the untreated would not be defined.

[93] The latter is a consequence of the discrete choice framework for schooling choices analyzed in the Heckman et al. (1998d) model. Recall our discussion in Section 3.4.

magnitude smaller than those reported in the literature estimating microeconometric treatment effects. The assumptions used to justify the LATE parameter in a microeconomic setting do not carry over to a general equilibrium framework. Policy changes, in general, induce two-way flows and violate the monotonicity – or one-way flow – assumption of LATE. Heckman et al. (1998e) extend the LATE concept to allow for the two-way flows induced by the policies. They present a more comprehensive approach to program evaluation by considering both the tax and benefit consequences of the program being evaluated and placing the analysis in a market setting. Their analysis demonstrates the possibilities of the general equilibrium approach and the limitations of the microeconomic "treatment effect" approach to policy evaluation.

9.3. Summary of general equilibrium approaches

Any policy with a large target population is likely to have general equilibrium impacts. Reliance on microeconomic treatment effect approaches to evaluate such policies produces potentially misleading estimates. Even reducing the Heckman et al. (1998e) estimates by a factor of three to account for learning about future price paths, instead of perfect foresight, produces a sizeable discrepancy between the microeconomic treatment effect estimates and the general equilibrium estimates. Their work and that of Davidson and Woodbury (1993) indicates that the costs of ignoring indirect effects may be substantial. In future evaluations of large scale programs, we urge the use of general equilibrium methods to produce more accurate assessments of the true impacts of the programs being evaluated and to produce a more reliable guide to the distributional impacts of policies.

The cost of this enhanced knowledge is the difficulty in assembling all of the behavioral parameters required to conduct a general equilibrium evaluation. From a long-run standpoint, these costs are worth incurring. Once a solid knowledge base is put in place, a more trustworthy framework for policy evaluation will be available, one that will offer an economically justified framework for accumulating evidence across studies and will motivate empirical research by microeconomists to provide better empirical foundations for general equilibrium policy analyses.

10. A survey of empirical findings

10.1. The objectives of program evaluations

The purpose of government training programs and other active labor market policies is to integrate unemployed and economically disadvantaged workers into the work force either by facilitating their job search, improving their work habits, or augmenting their human capital. In Section 3, we emphasized that program evaluators could assess the success of these programs by their impacts on a variety of outcomes, the choice of which depended on the objectives of policy makers. In practice, the outcomes of greatest interest to program evaluators and to policy makers who fund this research include participants'

labor market outcomes, such as their earnings, employment rates, transition rates out of unemployment and employment, wages, and use of unemployment insurance programs. Participants' non-labor market outcomes, such as their use of social assistance programs, educational attainment, criminal activity, and teen childbearing, are also scrutinized.

The most common outcomes of interest in US program evaluations are annual or quarterly earnings. Positive earnings impacts are often taken as synonymous with increased aggregate output and costs are often ignored. By contrast, in European evaluations the most common outcome of interest is employment. This emphasis reflects an emphasis on programs that reduce longterm unemployment.

Besides examining the impact of active labor market policies on participants' outcomes, another objective of program evaluations is to determine whether these policies constitute worthwhile social investments. The dominant approach followed in the program evaluation literature is to measure the net social benefit of these policies using the change in aggregate output attributable to the program (Heckman and Smith, 1998a). Evaluators estimate this change by subtracting the programs' costs from its discounted stream of benefits. These costs include the operating cost of the program, the cost of education and training expenditures, forgone earnings associated with participants' time in the program, and participants' out-of-pocket expenses for inputs such as transportation and child care.

In some cases, only the direct costs of these programs are likely significant in conventional cost-benefit analyses. The forgone earnings costs of participating in training are less important when evaluating JSA or short WE programs, or for programs targeted toward economically disadvantaged persons who are prone to long spells of unemployment. By contrast, these costs tend to be higher when evaluating a CT program in which individuals acquire skills off-the-job, and their participation in the program causes them to search less intensely for employment. Similarly, these costs are higher for programs serving adult males, especially prime-aged displaced workers, who have well established work histories and who are more likely to be employed in the absence of training.

In practice, conventional cost-benefit analyses usually do not account for several other costs that could reduce the net social benefit of employment and training programs. The first of these costs are the deadweight loss caused by raising taxes to finance training (Browning, 1987). The likely importance of these costs depends on the group being served. These costs should be higher for participants who are not receiving social assistance benefits. Often program evaluations report that the earnings impact of training is offset to some extent by a reduction in social assistance, so that participants' incomes may be little changed as a result of the program (see, e.g., Friedlander et al., 1985). This result implies that the deadweight loss associated with raising taxes to pay for training would be reduced to some extent because of savings in deadweight losses due to reduced taxes required to pay participants' future social assistance benefits.

A second cost usually unaccounted for in program evaluations is the value of participants' reduced leisure time (Greenberg, 1997). In principle, such costs depend on the shape of labor supply curves for different groups of participants. The value of participants' reduced leisure time may be especially significant for economically disadvantaged

women. If these women are the primary child care providers in their households, the social (as well as the private) cost associated with their time away from the home may be significant.

Finally, a third cost usually unaccounted for, especially in US program evaluations, is the cost associated with displacement of non-training participants (Hamermesh, 1971, 1993; Johnson, 1979). As discussed in Section 9, a potentially important policy parameter is the impact of the program on non-participants. If non-participants are displaced from jobs as a result of providing employment and training opportunities to participants, the program may have no impact on aggregate output. In the US, where a larger share of training dollars is spent on CT, the size of the programs compared to the economy is very small, and real wages have been relatively flexible, these costs are relatively small. In such instances, the estimated earnings impacts of the program may closely approximate the impact of the program on aggregate output.

By contrast, many European countries' active labor market polices include substantial expenditures on wage subsidies. These policies in the context of less flexible labor markets suggest that the cost of displacement, substitution and deadweight, as defined in Section 9, can be substantial. Evidence on this is given for the United Kingdom by Begg et al. (1991) and Dolton (1993), for the Netherlands by de Koning (1993) and for Sweden by Forslund and Krueger (1997). See Calmfors (1994) for a general survey.

The benefits from employment and training programs can come from several sources. By design, the discounted earnings impacts should be an important social benefit of most successful programs. In principle, other outcomes also could yield substantial social benefits. These outcomes include the value of output produced by trainees while in training, and the savings in administrative costs because of participants' reduced use of social welfare and of other education and training programs. Further, if improved employment prospects reduce asocial behaviors, society also may benefit from reduced expenditures on the criminal justice system, on substance abuse treatment centers, or on child welfare services. These latter benefits are potentially large for younger, less educated, training participants who are more inclined to engage in such asocial behavior (Mallar et al., 1982; LaLonde, 1995; Heckman and Smith, 1998a).

As shown by Table 18, the primary social benefit reported in most cost-benefit analyses of employment and training programs is the discounted earnings gains. Although this table surveys only a few analyses for economically disadvantaged women, these results are typical of those reported in other studies. Usually, these earnings benefits are one or two orders of magnitude larger than the other measured benefits of these programs. Because of the importance of earnings impacts for conventional cost-benefit analyses, it is important that analysts obtain credible and precise estimates of their magnitude.

The importance of estimated earnings impacts to cost-benefit analyses of employment and training programs highlights an important shortcoming of these analyses. As shown by the last row in Table 18, most program evaluations follow participants only for a couple of years following their entry into the program. Often the earnings impacts during this period

Table 18
Accounting of estimated social benefits and costs per treatment in selected social experiments evaluating employment and training services for female welfare applicants and recipients (1997 US dollars)[a]

	Program/main services provided NSW WE	San Diego CWEP/ JSA/WE	San Diego CWEP/ JSA Only	San Diego SWIM/ JSA/CT	Florida PI/ JSA	NJS All
Benefits						
Increased output from employment (includes earnings and fringe benefit impacts)	24486	3571	2457	2913	757	2066
From projected period only[b]	19084	2101	1163	0	298	0
Value of in-program output	12039	3280	−5	262	NA	NA
Reduced cost of using transfer programs	2160	131	82	53	130	NA
Reduced cost of using other programs (e.g., other education and training programs)	1619	85	74	NA	NA	NA
Costs						
Program operating costs, including JSA or WE	−13850	−968	−857	−866	−417	−1421[d]

Education and training costs	0	0	0	−360	−846	NA[d]
Forgone earnings and fringe benefits[c]	−2341	NA	NA	NA	NA	NA
Participant out-of-pocket expenses (e.g., transportation, child care, clothing costs)	−431	−24	–[e]	–	–	–
Value of reduced non-market time	–	–	–	–	–	–
Displacement of other workers	–	–	–	–	–	–
Deadweight loss from taxes to pay for programs	–	–	–	–	–	–
Net present value of benefits minus costs	21708	3123	1753	2003	−377	645
Observation period in years	2.25	1.00	1.16	5.00	2.00	2.50

[a] Sources: Kemper et al. (1981, Table IV.1, p. 100, Table IV.2, p. 106, Table IV.6, p. 121); Goldman et al. (1986, p. 139, Table 5.4, p. 153, Table 5.8, p. 166); Friedlander and Hamilton (1993, Table 5.1, p. 57); Kemple et al. (1995, Table 7.3, p. 174, Table 7.5, p. 177); Orr et al. (1994, Exhibit 6.2, p. 162).

[b] Projected earnings are based on earnings impacts during the last four quarters of the observation period and are discounted at a rate of 5% per year. Subsequent earnings impacts were assumed to depreciate at a rate of 25% per year.

[d] The −1421 figure for the NJS includes both program operating costs and education and training costs.

[c] For all but the first column, the social costs associated with forgone earnings are embodied in the estimates of the increased output from employment. In these cases this measure is net of the forgone earnings costs of the program.

[e] A "–" denotes costs that were not estimated in the indicated study.

are insufficient to justify the programs' costs. Cost-benefit analyses in this literature customarily project the earnings impacts obtained during the observation period into the future, sometimes for as long as participants' expected working life, and then discount these projected impacts at rates ranging from 0 to 15% (see, e.g., Kemper et al., 1981, pp. 174–177, Table VIII.2). In addition, evaluators sometimes allow these projected impacts to decay through time (see, e.g., the references in Table 18).

As shown by the third row of Table 18, the projected earnings gains can constitute a significant portion of the total earnings gains associated with the program. In the most extreme case, more than three-fourths of the earnings impact used in the cost-benefit analysis of the NSW Demonstration was based on out-of-sample projections. Because the estimated benefits from the reduced use of other social programs by NSW treatments also were based on similar projections, the net social benefit of the NSW Demonstration during the first 27 months is actually negative. This evidence underscores the importance of funding the collection of data that enable longterm evaluations of employment and training programs.

For evaluations that look only at post-program earnings impacts, another potential source of benefits from active labor market policies is the value of the output produced by participants while they were in the program.[94] As shown by the first column of Table 18, this benefit constituted a significant fraction of the total benefit from the NSW program. This result is expected in programs that provide WE compared with those that provide JSA or CT. When training consists of a subsidized job in the public or non-profit sector, evaluators assume that the work performed by participants is valuable to society. Because the NSW program provided relatively longterm WE to a large percentage of participants, the value of in-program output is large compared to other programs. In contrast, the WE in the San Diego CWEP program shown in the second column lasted only a few weeks and was provided to only a small fraction of participants. As a result the value of in-program output was small.

To demonstrate the sensitivity of conventional cost-benefit analyses to assumptions about the costs and benefits of employment and training programs, we reexamine the net social benefits of the WE provided in the NSW Demonstration and of JTPA services provided in the NJS. Since the "final reports" for these two studies were published, there have been two subsequent studies that have followed participants for up to 8 and 5 years, respectively (Couch, 1992; US General Accounting Office, 1996). Both of these studies indicate that the positive shortterm earnings impacts originally reported for adult women in the NSW Demonstration and adults in the NJS persisted, whereas neither program had a significant short- or longer-term impact on youths' earnings.

As shown by Table 19, the estimated net social benefit of treatments' access to WE in the NSW Demonstration are negative for youths, but are sometimes positive for adult

[94] When the in-program period is included in the estimation of program impacts, and participants are paid a market wage, the value of in-program output is implicitly included in the impact estimate because it is reflected in the participants' earnings.

Table 19
Net social returns and internal rates of return: National Supported Work Demonstration (impacts and costs in 1978 US dollars)[a]

Benefit duration	Welfare cost of taxation	AFDC Women IRR[b]	Youth IRR	Annual discount rate	AFDC Women net social benefit	Youth net social benefit
3 years	0.00	<0	<0	0.00	−2152	−1528
				0.05	−2167	−1541
				0.10	−2180	−1553
	0.50	<0	<0	0.00	−3489	−2406
				0.05	−3504	−2419
				0.10	−3517	−2430
	1.00	<0	<0	0.00	−4826	−3283
				0.05	−4841	−3296
				0.10	−4854	−3308
8 years	0.00	0.005	<0	0.00	54	−1463
				0.05	−428	−1482
				0.10	−789	−1499
	0.50	<0	<0	0.00	−1283	−2341
				0.05	−1765	−2359
				0.10	−2126	−2377
	1.00	<0	<0	0.00	−2620	−3218
				0.05	−3102	−3237
				0.10	−3463	−3254
Indefinite	0.00	0.136	<0	0.00	NA[c]	NA
				0.05	4648	−1942
				0.10	961	−1658
	0.50	0.091	<0	0.00	NA	NA
				0.05	3311	−2820
				0.10	−376	−2535
	1.00	0.068	<0	0.00	NA	NA
				0.05	1974	−3697
				0.10	−1713	−3413

[a] Sources: Impact estimates are taken from Couch (1992). Cost estimates are taken from Kemper et al. (1984, Table 8.6). Estimates of the welfare cost of taxation fall within the range given in Browning (1987).

[b] **IRR**, internal rate of return, the rate of return at which the discounted benefits from the program equal the current costs. Welfare costs of taxation are in dollars of welfare loss per tax dollar.

[c] NA indicates that net social benefits equal positive or negative infinity due to the absence of discounting.

women. However, the table reveals that these estimates are sensitive to the duration of the earnings impacts, the discount rate used in the analysis, and whether the analysis takes into account the deadweight losses associated with the taxes that finance the program. Although the earnings impacts for adult women are positive during the first 3 years,

these impacts by themselves are insufficient to generate positive net social benefits from the program.[95] The importance of the followup study by Couch (1992) is seen in the estimates of the net social benefits from the program when the benefits last for 8 years or indefinitely. In the latter case, estimates based on a variety of plausible assumptions about the appropriate discount rate and the deadweight loss associated with taxes all imply that the estimated net social benefit from the program is positive for AFDC women.

A useful metric for comparing the net social benefits of different active labor market policies to other investments is their internal rate of return (IRR). This measure is the discount rate for which the discounted stream of benefits from the program equals its costs. The IRR allows a comparison between alternative investment projects using a common metric. As shown by the middle columns of Table 19, if we assume that the deadweight loss associated with taxes used to finance the program are 50% or more, the IRR for WE targeted toward adult women is negative for benefit durations of 8 years or less. If the earnings impacts persist indefinitely, the IRR is 9.1%. Thus, for adult women in the NSW, the overall net benefit calculations still depend on projections of earnings gains outside the available data.

Comparing Tables 20 and 21, the cost-benefit analyses indicate that JTPA services generated a substantial net social benefit when targeted toward adults, but none when targeted toward youths. As with the NSW Demonstration, these estimated net social benefits are sensitive to the assumptions underlying the analysis. In the absence of a longterm followup study, we would be less confident about whether JTPA constituted a worthwhile social investment. However, as a result of the followup study, we are more confident that after 5 years the net social benefit per treatment group member ranges from 600 to 2000 and that the IRR are very large. Further, if these gains were to persist indefinitely, it would appear that the JTPA services provided adults in the NJS constituted an extraordinarily successful public investment. By contrast, as shown by Table 21, estimates based on short- and medium-term earnings impacts indicate that JTPA services targeted toward youths constituted a poor social investment. As shown by the last rows of the table, projections of these impacts into the future produce the only positive net social returns for this group. However this result is very tenuous because these projections are based on point estimates for the fifth (followup) year that are not statistically significant.

10.2. The impact of government programs on labor market outcomes

Credible cost-benefit analyses of employment and training programs depend on credible estimates of the costs and benefits of these programs. Because labor market outcomes appear to constitute such an important source of the social benefits from these programs

[95] As we discussed in Section 5, the impact estimates for the NSW Demonstration differ depending on whether the survey earnings data or the administrative earnings data are used, with the estimates based on the survey measures showing a larger positive impact. Because the only longterm followup impact estimates are based on the administrative data (Couch, 1992), we use them in constructing the estimates in Table 19.

Table 20
Net social returns and internal rates of return: National JTPA Study – Adults (impacts and costs in nominal US dollars)[a]

Benefit duration	Welfare cost of taxation	Adult women IRR[b]	Adult men IRR	Annual discount rate	Adult women net social benefit	Adult men net social benefit
3 years	0.00	1.390	>2	0.00	863	1097
				0.05	778	1017
				0.10	702	948
	0.50	0.416	1.787	0.00	485	844
				0.05	400	765
				0.10	324	696
	1.00	0.064	0.689	0.00	107	592
				0.05	22	513
				0.10	−54	443
5 years	0.00	1.610	>2	0.00	1822	1979
				0.05	1589	1766
				0.10	1395	1589
	0.50	0.693	>2	0.00	1443	1726
				0.05	1211	1514
				0.10	1017	1336
	1.00	0.362	0.960	0.00	1065	1474
				0.05	833	1261
				0.10	638	1084
Indefinite	0.00	1.620	>2	0.00	NA[c]	NA
				0.05	7889	6859
				0.10	3891	3607
	0.50	0.738	>2	0.00	NA	NA
				0.05	7510	6607
				0.10	3513	3354
	1.00	0.455	0.985	0.00	NA	NA
				0.05	7132	6354
				0.10	3134	3102

[a] Impact estimates are taken from US General Accounting Office (1996). Cost estimates are taken from Orr et al. (1994). Estimates of the welfare cost of taxation fall within the range given in Browning (1987).

[b] IRR, internal rate of return, the rate of return at which the discounted benefits from the program equal the current costs. Welfare costs of taxation are in dollars of welfare loss per tax dollar.

[c] NA indicates that net social benefits equal positive or negative infinity due to the absence of discounting.

and because these outcomes are relatively easily measured, they are the focus of much of the evaluation literature. There is much less emphasis in the literature on the impact of these programs on non-labor market outcomes.

There have been many surveys of the impact of US programs on labor market outcomes, especially on participants' employment rates and earnings (see, e.g., Perry et al., 1975;

Table 21
Net social returns and internal rates of return: National JTPA Study – Youth (impacts and costs in nominal US dollars)[a]

Benefit duration	Welfare cost of taxation	Female youth IRR[b]	Male youth IRR	Annual discount rate	Female youth net social benefit	Male youth net social benefit
3 years	0.00	<0	<0	0.00	−982	−2196
				0.05	−979	−2145
				0.10	−976	−2101
	0.50	<0	<0	0.00	−1413	−2849
				0.05	−1410	−2798
				0.10	−1407	−2754
	1.00	<0	<0	0.00	−1844	−3502
				0.05	−1841	−3451
				0.10	−1838	−3407
5 years	0.00	<0	<0	0.00	−434	−1158
				0.05	−515	−1281
				0.10	−580	−2027
	0.50	<0	<0	0.00	−865	−1811
				0.05	−946	−1934
				0.10	−1011	−2027
	1.00	<0	<0	0.00	−1296	−2464
				0.05	−1377	−2587
				0.10	−1442	−2680
Indefinite	0.00	0.163	0.163	0.00	NA[c]	NA
				0.05	2995	10880
				0.10	811	3444
	0.50	0.122	0.122	0.00	NA	NA
				0.05	2564	10227
				0.10	380	2791
	1.00	0.098	0.098	0.00	NA	NA
				0.05	2133	9573
				0.10	−51	2138

[a] Impact estimates are taken from US General Accounting Office (1996). Cost estimates are taken from Orr et al. (1994). Estimates of the welfare cost of taxation fall within the range given in Browning (1987).

[b] IRR, internal rate of return, the rate of return at which the discounted benefits from the program equal the current costs. Welfare costs of taxation are in dollars of welfare loss per tax dollar.

[c] NA indicates that net social benefits equal positive or negative infinity due to the absence of discounting.

Grossman et al., 1985; Bassi and Ashenfelter, 1986; Barnow, 1987; Gueron, 1990; LaLonde, 1995; Friedlander et al., 1997). By contrast, there are few surveys of the impacts of these programs operated outside the US (see, e.g., Bradley, 1994; Fay, 1996). Consequently, to address this imbalance we devote a substantial portion of this section to

summarizing what has been learned from evaluations of European programs, and how these studies compare to US evaluations.

Before surveying the results from these evaluations it is helpful to consider what kind of impact on earnings we should expect from public sector employment and training programs. If we believe that the objective of these programs is to augment human capital, the literature on education and earnings provides a useful starting point. This literature indicates that an additional year of schooling is associated with approximately a 10% increase in the typical worker's earnings (Ashenfelter and Rouse, 1995). In many countries, the return to schooling is smaller, as it also has been in the past in the United States. The cost of a year of education includes direct instructional expenditures, any forgone earnings, and other inputs from the family and the community. Formal schooling is usually more intensive and costly than public sector employment and training programs. As a result, it would be surprising if such programs, which usually last far less than a year, consistently led to larger increases in earnings than an additional year of schooling. By analogy, the relatively few programs that are more intensive and costly than a year of schooling should generate larger earnings gains (Heckman et al., 1993; LaLonde, 1995). Accordingly, a training program costing several hundred dollars or even a few thousand dollars per participant would likely lead to annual earnings gains of at most several hundred dollars. Earnings gains much larger than this would suggest that these programs generate large social returns compared to formal schooling and to other investments in general. In this vein, the results in Table 20 showing very high estimated internal rates of return from JTPA are unexpectedly large.

The evidence both from North American and European studies indicates that government employment and training programs have at best a modest positive impact on adult earnings. Further, when longer term followup data are available, these gains do not always persist. The evidence suggests that the gains, when they occur, are more likely the result of an increased probability of employment than of increased wages. Indeed, the case for these programs increasing participants' subsequent hourly wages remains weak. The finding that earnings gains are in large measure the result of increased employment rates raises the question of how active labor market polices affect non-participants through displacement. In the US especially, this issue has received relatively little empirical attention (but, see Davidson and Woodbury, 1993, 1995).

Among youths, the evidence is mixed between the two continents. In the US, studies consistently report that these programs have no impact (or sometimes even a negative impact) on youths' earnings. By contrast, in Europe, studies of less economically disadvantaged youths find that these programs sometimes substantially raise employment rates, because they raise transition rates out of unemployment. At the same time, however, other studies (sometimes of participants in the same program) report no effect on employment. Such results suggest either that there is substantial heterogeneity in impacts among cohorts or that these impacts possess the same sensitivity to econometric specification that we documented for the US CETA studies. In any case, as with adults there is little evidence that these programs raise wages.

Part of the reason that there is little evidence on the relation between government training programs and wages has to do with the quality of data available for program evaluations. Hourly wage data are unavailable in many program evaluations, especially those conducted in the United States. Further, when these measures are available, the sample sizes often are too small to estimate wage impacts with precision. From a human capital perspective, the wage gains associated with these programs should be small. The experiences with the NSW Demonstration and NJS illustrate this problem. In the former program the wages of adult women in the treatment group were approximately 8% higher than those in the control group; in the later program wages of adults in the treatment group were 2–3%% higher than those in the control group (Masters and Maynard, 1981; Orr et al., 1994). Neither of these impacts were statistically significant at conventional levels. However, given the costs of these programs, these point estimates are not surprising. Indeed, they compare favorably to estimates of the wage impacts associated with a year of community college schooling in the US (Kane and Rouse, 1993).

At the same time, this characterization of the empirical results from the program evaluation literature masks substantial heterogeneity in the estimated impacts which vary widely among programs, among field offices and among different demographic and skill groups. In many instances, the evidence suggests that training either had no effect or may have lowered earnings, while in other cases the impacts are so large that programs such as JTPA appear to generate substantial internal (and private) rates of return. Indeed, for economically disadvantaged adult women residing in the US, a case can be made that these programs consistently have been a productive social investment, whose returns are larger than those from formal schooling. For other groups this conclusion clearly does not hold. In particular, there appears to be a weak tendency in the literature suggesting that the earnings impacts and the net social returns from many active labor market policies, particularly those that provide training, are smaller for the least skilled participants.

10.3. The findings from US social experiments

As explained in Section 5, an unusual characteristic of the empirical literature on active labor market policies is that it includes a relatively large number of both experimental and non-experimental studies. However, because treatment non-participation and control group substitution are often substantial, the parameter measured in experimental studies is the effect of the "intention to treat" and not the impact of "treatment on the treated" (Heckman et al., 1998f). Dropping out similarly afflicts non-experimental studies, and contamination bias is the counterpart to control group substitution. Accordingly, although the estimates reported in the experimental literature are usually thought to be different from those in the non-experimental literature, it is easy to exaggerate the differences. Nonetheless, because the estimates reported in both literatures do not adjust for these biases, and because the incidence of the various biases may differ between experimental and non-experimental studies, it is likely that different parameters are

estimated in these diverse literatures. For these reasons, we survey the two literatures separately.

Provided the assumptions discussed in Section 5 hold, social experiments yield easily computed and widely understood estimates of the "the intention to treat" on the treatments' outcomes. As shown by Table 22, collectively, the US experimental evaluations provide some compelling evidence that the opportunity to receive these services sometimes can improve participants' employment prospects and that the resources spent on these services can pass a standard cost-benefit test. The most consistent evidence in this regard is found for adult women.[96] As shown by Table 22, the earnings gains received by adult women assigned to the treatment group are (i) usually modest in size ranging from a few hundred dollars to more than one thousand dollars, annually, (ii) often persist at least for several years without signs of decay, (iii) arise from a variety of intended treatments, and (iv) sometimes appear to be remarkably cost effective, at least before the deadweight costs of taxation, displacement and substitution effects are taken into account. Further, although the opportunity to receive job search assistance appears to be the most cost-effective service in the sense that it has the highest IRR, more expensive WE and training programs result in larger absolute earnings gains.

Because of substantial treatment non-participation and control group substitution, the impact of these services on those who actually received them is generally larger than indicated by the experimental estimates reported in Table 22. The exceptions are the NSW and AFDC Homemaker-Health Care demonstrations. As explained in Section 5, the NSW provided relatively longterm WE. The AFDC Homemaker Demonstrations trained economically disadvantaged women to provide in-home care to the disabled and the elderly (Bell and Reesman, 1987). Participation rates in these relatively expensive treatments were high and similar services were generally unavailable to the controls (Masters and Maynard, 1981, p. 148; Bell and Reesman, 1987, p. 14). Therefore, in these two studies the experimental impacts can reasonably be interpreted as approximating the impact of the "treatment on treated."

As suggested by the number of studies surveyed in Table 22, there have been fewer experimental evaluations of the impacts of employment and training programs for adult men and especially for youths. As a result, the evidence based on social experiments is more fragmented. Nevertheless, the evidence suggests that programs that offer training can raise the earnings of economically disadvantaged adult males, but programs that focus on

[96] In keeping with the emphasis of US policy on reducing reliance on social assistance, most social experiments have tested the impact of employment and training services on individuals who were applying for or receiving social assistance or welfare (AFDC). The number of these experiments proliferated during the 1980s after the federal government authorized states to operate as demonstration projects community work experience programs (CWEP) for their welfare population. In several states, officials implemented an experimental design in a few welfare offices by mandating that only a random sample of the eligible population participate in JSA, CWEP, or other employment related activities (Goldman et al., 1986). Because the vast majority of social assistance recipients are single female household heads, this has meant that most of the experimental evidence relates to economically disadvantaged adult women. These experimental results were influential in shaping US welfare policy during the late 1980s (Greenberg and Wiseman, 1992).

JSA or WE appear to be ineffective or sometimes worse. Earnings impacts of the San Diego CWEP program, the Baltimore Options program, and the NSW Demonstration were small or negative for disadvantaged adult men. By contrast, the impacts reported in programs that offered training opportunities, San Diego-SWIM program, GAIN, and the NJS, were larger and statistically significant. In particular, the NJS found that economically disadvantaged adult men experienced earnings gains similar to those achieved by adult women (Orr et al., 1994, p. 82).

The evidence from experimental evaluations for youths is not encouraging. As shown by the last panel of Table 22, the results suggest that the array of services currently offered do little to raise youth employment and earnings. For example, the prolonged WE provided to disadvantaged high school dropouts in the NSW Demonstration had no effect on their earnings during the 8 years after the treatment was offered (Couch, 1992). Similarly, the JOBSTART demonstration, which provided disadvantaged youths with services similar to those offered by the comprehensive Job Corps program, but without the residential living centers, did not generate significantly higher earnings for the treatments during the 4-year followup period (Cave et al., 1993). Finally, the NJS finds no evidence that youth served by JTPA benefit from its relatively low cost training services. In fact the shortterm point estimates for the males were actually negative.

Another finding highlighted in Table 22 is the correspondence between earnings impacts and employment impacts. In most cases large earnings impacts are accompanied by significant impacts on employment rates. Moreover, in most of these studies analysts measure employment rates at the quarterly level and information on hours of work are unavailable. When such measures are available, hours impacts also can be a significant source of earnings gains (see, e.g., the NSW Demonstration, Hollister et al., 1984). Indeed, there are only two cases in the table for which the long-run earnings impacts are significant, but not the impact on employment rates. This evidence underscores the concern that because access to government employment and training programs raises earnings through higher employment rates, displacement of non-participants may mitigate the net social benefits reported for these treatments in conventional cost-benefit analyses.

The experimental impacts reported in Table 23 indicate that the impact of the opportunity to participate in particular employment and training services varies substantially among demographic groups. The WE services provided in the NSW demonstration were effective for adult women, but not youths; the WE provided in the San Diego CWEP demonstration was more effective for female welfare applicants than for their male counterparts. The JSA and training experiences provided in the San Diego SWIM demonstration also had a larger impact on women than on men.[97] Finally, the NJS reported striking differences between the impact of JTPA services on adults and youths. These results raise the issue of the importance of impact heterogeneity in this literature.

[97] These differences in experimental impacts are not the result of differing participation rates in the programs by women and men. In the San Diego SWIM Demonstration participation rates in programs services among female (i.e., AFDC-FG) and male (AFDC-U) treatments were nearly the same. Male controls were less likely than female controls to obtain the same services elsewhere (see Freidlander and Hamilton (1993, p. 22, Table 3.1).

Table 22
Impacts from US social experiments evaluating employment and training programs (difference between the treatments' and controls' mean employment rates and earnings in 1997 US dollars)[a]

Demographic group/ services tested/study	Average net costs[b]	Impacts on employment rates and earnings			
		Employment rate last quarter[c]	Earnings[d] Year 1/2	Earnings Year 3/4/5	% of control earnings
A. Economically disadvantaged females					
Job search assistance					
Arkansas WORK	244	6.2*	339*	487*	31
Louisville (WIN-1)	206	5.3*	425*	643*	18
Cook County, IL	231	1.2	12	NA	1
Louisville (WIN-2)	340	14.2*	679*	NA	43
San Diego – CWEP	891	−0.7	402*	NA	8
Food Stamp E & T	180	−2.5	−90	NA	−3
Minnesota – MFIP[e]	NA	14.5*	921*	NA	30
Job search assistance and work experience					
West Virginia	388	−1.0	25	NA	4
Virginia ES	631	4.6*	106	387*	11
San Diego – CWEP	690	3.8*	1120*	NA	23
Baltimore Options	1407	0.4	231	764*	17
Job search assistance and CT or OJT services					
Maine TOPS	2972	1.1	433*	1720*	36
San Diego SWIM	964	0.3	509*	180	15
New Jersey	1165	NA	874*	NA	14
GAIN (JOBS):	3757	5.9[f]	339*	740*	25
Alameda (Oakland)	6036	6.0*	266	901*	37
Los Angeles	6356	1.9	−5	178	9
Riverside	1753	7.5*	1173*	1176*	40
San Diego	2099	2.7*	445*	830*	23
MFSP San Jose (CET)	5132	8.6*	1470*	NA	25
MFSP Other Sites	4525	1.2	400	NA	6

Table 22 (*continued*)

Demographic group/ services tested/study	Average net costs[b]	Impacts on employment rates and earnings			
		Employment rate last quarter[c]	Earnings[d] Year 1/2	Earnings Year 3/4/5	% of control earnings
Florida PI (JOBS)	1339	0.4	93	NA	3
Work experience and training					
National Supported Work	8614	7.1	657	1062	43
AFDC Homemaker	8371	NA	2135*	NA	NA
NJS (JTPA)	1028	NA	691*	441*	7
Recommended for CT	1690	NA	359	NA	NA
Recommended for OJT	643	NA	747*	NA	NA
B. Economically disadvantaged males					
Job search assistance					
San Diego – CWEP	931	0	−325	NA	5
Job search assistance and work experience					
San Diego – CWEP	597	−1.2	−461*	NA	−8
West Virginia	210	NA	−234	NA	−6
Baltimore	1014	NA	−2564	NA	−27
Job search assistance and CT or OJT services					
San Diego – SWIM	747	2.1	704*	−305	−5
GAIN (JOBS):	3202	4.5[f]	489*	413*	11
Alameda (Oakland)	NA	7.9	69	615	18
Los Angeles	4885	9.9*	330*	1033*	22
Riverside	2361	3.8*	970*	389	10
San Diego	2251	−1.3	308	−300	−6
Work experience and training					
NSW-Ex-Offenders	10797	−0.9	100	NA	4

NSW-Ex-Addicts	12150	17.2*	86	706[g]	32
NJS (JTPA)	660	NA	668*	357	7
Recommended for CT	900	NA	895	NA	NA
Recommended for OJT	753	NA	1249*	NA	NA
C. Economically disadvantaged youths[h]					
Work experience and training					
National Supported Work	9314	0.3	−79	−79	−4
JOBSTART	6403	−0.9	−721	523	8
NJS (JTPA)					
Females	1116	NA	133	246	4
Males[i]	1731	NA	−553	852	11

[a] Sources: LaLonde (1995, p. 159); Greenberg and Wiseman (1992, pp. 52–53, 56); Gueron (1990, pp. 92–93); Goldman et al. (1986, pp. 54, 102, 241); Friedlander and Hamilton (1993, pp. 57, 117–118, 133–134); Friedlander et al. (1995, p. xx); Riccio et al. (1994, pp. 254, 267, 316–328, 350–361); Hollister et al. (1984, pp. 148, 181, Table 6.2); Orr et al. (1994, pp. 64, 65, 82, 104, 131, 151, 162, 165, 166); US General Accounting Office (1996, pp. 20–21, Tables ii.1, ii.2, ii.3, and ii.4); Puma and Burstein (1994, pp. 322–23, 325); Knox et al. (1997, Table 5.3). An asterisk indicates that the impact is statistically significant at the 10% level.

[b] Average net costs are the incremental costs of providing services to the members of the treatment group.

[c] "Employment rate last quarter" refers to the difference between treatments' and controls' employment rates during the last quarter of the followup period for which data was available.

[d] The earnings impacts are annual (or annualized) differences between the treatments' and controls' mean earnings during the first or second year (Year 1/2) and during the third, fourth, or fifth year (Year 3/4/5).

[e] Figures are for longterm welfare recipients only. Two other components of this program included both threat of sanctions and financial incentives for welfare recipients to find work.

[f] Measure of ever employed during the last year of followup instead of the last quarter.

[g] Standard error not available.

[h] These studies also examined the impacts on arrests. In the NSW demonstration, the percentage of treatments ever arrested during the study's first 27 months was 8.8 percentage points (22%) less than the controls (Maynard, 1980, p. 138, Table VI.5); In the JOBSTART demonstration the number of treatments' arrested was 2.6 percentage points (21%) lower than the number of controls during the first year of the study. During the 4-year study the percentage of arrests among both experimental groups was the same (Cave et al., 1993, p. 195, Table 6.7); In the NJS, the percentage of treatments arrested was higher than for the controls; the percentage of males with no prior arrests (since age 16) before the study who were subsequently arrested was 7.1 percentage points (38%) larger for the treatments than for the controls (Orr et al., 1994, p. 117, Exhibit 4.22).

[i] Sample of male non-arrestees.

Just as this impact heterogeneity is found among different demographic groups, it also is often found among different sites in the same study. When experimental impact estimates for the same program are available for different sites, it is common to find that the impacts vary among sites. For example, as shown by Table 23, the results from the GAIN program and Minority Female Single Parent Demonstration (MFSP) reveal substantial variation in impacts among sites. Similar variation in experimental impacts also is reported among the 10 sites in the NSW Demonstration and the 16 sites in the NJS (see Maynard, 1980, p. 83; Masters and Maynard, 1981, p. 85; Heckman and Smith, 1998b). At the very least, this evidence of heterogeneity in impacts among sites raises the question of the external validity of these evaluations, i.e., whether their results can be extended to other settings. For policy purposes it is important to know whether the differences in site impacts arise from differences in the skills of program operators and trainers, program organization, or the characteristics of those who are served.

The experimental evidence can shed some light on how heterogenous the impacts are among those served by these programs. An important question in this regard is whether government training programs generate different returns for participants depending on their observed and (to the econometrican) unobserved skills. If returns are smaller for the least skilled, then policy makers would be faced with the difficult question of whether to reallocate expenditures toward less "needy" participants. In 1981, US policy makers in fact made the opposite decision when they directed that employment and training expenditures be targeted to a more economically disadvantaged population (Barnow, 1987). An important policy question is whether this decision improved or worsened the returns from these social programs.

Neither the experimental nor the non-experimental evidence provides a clear answer to the question of whether the impacts of these programs vary with participants' skills. But the experimental evidence does suggest that the least able participants among the low-skilled populations served by these programs benefit the least from them, especially when the programs provide CT and OJT opportunities. To illustrate these points, Table 23 presents the experimental impacts by the prior skills of participants for several social experiments. The measures of skill differ among studies, but as indicated by the controls' earnings during the followup period, these differing measures of skill correctly identify individuals likely to perform poorly in the labor market. In the GAIN and NJS studies more skilled persons benefited more from access to the program's services than did less skilled persons. However, as the table demonstrates, in some programs, such as the NSW and the San Diego CWEP Demonstrations, the least skilled experienced larger gains. Significantly, these programs provided treatments with WE. As explained in Section 2, the purpose of this service is to provide a job experience to individuals with poor employment histories so that they can develop acceptable "work habits." By design, therefore, it might be expected that this service would provide greater benefit to less skilled participants than to more skilled participants who already possess such skills.

Table 23
Experimental impacts of employment and training programs on earnings by prior skills of participants (impacts in nominal US dollars)[a]

Evaluation/total followup period in years/skill measure	Controls' earnings	Impact on earnings[b]	Percentage impact
A. *Economically disadvantaged female household heads*			
NSW/2.25 years			
9–11 years of school	324[c]	181*	52
HS Graduate	633[c]	72	11
San Diego CWEP/1.5 years			
Not employed during prior year	1474	1066*	72
Employed during prior year	4640	347	7
San Diego SWIM/5 years			
HS Drop-out	8783	1654	19
HS Graduate	18135	2405*	13
Florida Project Independence/2 years			
Never employed during prior 36 months	2117	318	15
HS Drop-out and worked 12/36 months	2904	209	7
HS Graduate and worked 12/36 months	6538	314	5
California GAIN/3 years			
a. Alameda Co. (Oakland):			
Assessed to need basic education	3826	610	16
Does not require basic education	8142	2947*	36
b. Los Angeles:			
Assessed to need basic education	3809	107	3
Does not require basic education	8142	1147*	14
c. Riverside:			
Assessed to need basic education	4408	2595*	59
Does not require basic education	9206	3950*	43

Table 23 (*continued*)

Evaluation/total followup period in years/skill measure	Controls' earnings	Impact on earnings[b]	Percentage impact
d. San Diego:			
Assessed to need basic education	5837	572	10
Does not require basic education	11026	3040*	28
Minority Female Single Parent Demonstrations/1 year			
a. Atlanta, Georgia - AUL:			
HS Drop-out	3967	576	15
HS Graduate	5948	−280	−5
b. San Jose, California - CET:			
HS Drop-out	4656	1068*	23
HS Graduate	5364	1368*	26
c. Providence, Rhode Island - OIC:			
HS Drop-out	3272	408	13
HS Graduate	4608	72	2
National JTPA Study/2.5 years			
HS Drop-out	9379	878	9
HS Graduate	13484	1152*	8
Received welfare for >2 years	8056	2255*	28
Never received welfare	14513	563	4
Never worked	6887	788	11
Earned <4 in last job	10979	943	9
Earned >4 in last job	14528	1626*	11
B. *Economically disadvantaged male household heads*			
NSW-Ex-addicts/3 years			
9–11 years of school	442[c]	142	32
HS Graduate	458[c]	320*	70
NSW-Ex-offenders/3 years			
9–11 years of school	596[c]	95	16
HS Graduate	622[c]	126	20
San Diego CWEP - JSA only/1.5 years			
Received welfare for >2 years	6911	1187	17

Table 23 (*continued*)

Evaluation/total followup period in years/skill measure	Controls' earnings	Impact on earnings[b]	Percentage impact
Never received welfare	7487	−364	−5
San Diego CWEP - JSA/WE/1.5 years			
Received welfare for >2 years	5724	1398	24
Never received welfare	7852	−280	−4
San Diego SWIM/5 years			
HS Drop-out	19329	−679	−4
HS Graduate	24645	3041	12
California GAIN/3 years			
a. Riverside:			
Assessed to need basic education	9398	555	6
Does not require basic education	11274	3461*	31
b. San Diego:			
Assessed to need basic education	5837	−515	−5
Does not require basic education	11026	1453	10
National JTPA Study/2.5 years			
HS Drop-out	14520	1353	9
HS Graduate	20018	918	4
Never worked	14368	−2104	−15
Earned <4 in last job	14268	245	2
Earned >4 in last job	19353	1647*	9
C. *Economically disadvantaged male youths*			
JOBSTART/4 years			
Not employed during prior year	20164	−1893	−9
Employed during prior year	24729	707	3
Arrested since age 16	20344	1553*	8
Not Arrested since age 16	23183	−921	−4
National JTPA Study/2.5 years (non-arrestees)			
HS Drop-out	14394	−1064	9

Table 23 (*continued*)

Evaluation/total followup period in years/skill measure	Controls' earnings	Impact on earnings[b]	Percentage impact
HS Graduate	19605	−484	4
Never worked	11052	587	5
Earned <4 in last job	16143	−1198	−7
Earned >4 in last job	19056	−1727	−9

[a] Sources: NSW: Masters and Maynard (1981, pp. 89–90); Hollister et al. (1984, pp. 154, 183); San Diego CWEP: Goldman et al. (1986, pp. 92, 126); San Diego SWIM: Friedlander and Hamilton (1993, pp. xxix and xxxi); Florida Project Independence: Kemple et al. (1995, p. 136); California GAIN: Riccio et al. (1994, pp. 137–138, 217–218); Minority Female Single Parent Demonstration: Rangarajan et al. (1992, Volume IV, pp. 37–41); National JTPA Study: Orr et al. (1994, pp. 135–137, 154); JOBSTART: Cave et al. (1993, pp. 156–163). HS, high school. An asterisk indicates that the impact is significant at the 10% level.

[b] Earnings impacts are the difference between treatments' and controls' nominal earnings during the entire followup period given in years next to the name of the program.

[c] Subgroup impacts in the NSW studies in Masters and Maynard (1981) and Hollister et al. (1984) are reported in terms of monthly hours. The figures in the table refer to the period during the last 9 months followed in the study multiplied by 9.

10.4. The findings from non-experimental evaluations of US programs

The experimental evaluations provide evidence that the opportunity to participate in employment and training programs (i) can improve the employment prospects of low skilled persons, and (ii) has markedly varying impacts on different demographic and skill groups. Non-experimental evaluations more often estimate the treatment on the treated parameter although partial participation and dropping out are an important part of ongoing programs as well (Heckman et al., 1998f). Patterns have emerged from these studies that are consistent with and reinforce the findings from the experimental literature.

These patterns exist despite the controversy about the sensitivity in non-experimental estimates and its implications for policy analysis, suggesting that the problems raised by the proponents of the experimental method may be exaggerated. As discussed earlier, the most striking result of non-experimental evaluations of US employment and training programs is the variability in the estimated impacts of training. Not only do the effects vary among different cohorts, but even when program evaluators assess the same cohort, they often arrive at substantially different estimates of the training effect. This sensitivity is one of the most important lessons from this literature and, as we discuss below, it is a lesson that emerges to some extent from the European experience as well. A dramatic illustration of this assertion is the evaluation of the US CETA program. As shown by Table 24, the impact estimates from six evaluations of the 1976 CETA cohort range from −$1553 to $1638 for male participants and from $24 to $2669 for female participants. Not surprisingly, one group of evaluators involved in these studies concluded that

Table 24
The impact of US Federal Government employment and training programs on participants' earnings (increase in post-program annual earnings in 1997 US dollars)[a]

Study	Training cohort[b]	Men[c] (whites/minorities)	Women (whites/minorities)
A. *Non-experimental estimates for economically disadvantaged adult participants*			
Ashenfelter (1978)	1964 MDTA	910/631	2111/1868
Kiefer (1979)	1969 MDTA	−2026/−2244	1905/2621
Gay and Borus (1980)	1969–1972 MDTA	152/161	1373/377
Cooley et al. (1979)	1969–1971 MDTA	1395	2038
Westat (1984)	1976 CETA	−12/−255	983/801
Bassi (1983)	1976 CETA	61/−1055	1286/2669
Dickinson et al. (1986)	1976 CETA	−1553	24
Geraci (1984)	1976 CETA	0	2026
Bloom/McLaughlin (1982)	1976 CETA	364	1844
Ashenfelter/Card (1985)	1976 CETA	1638	2220
Dickinson et al. (1986)	1/76–6/76 CETA	−1031	546
Westat (1984)	1977 CETA	1128/1480	1201/1711
Bassi et al. (1984)	Welfare		
	1977 CETA	1419/−231	2014/1529
Bassi et al. (1984)	Non-welfare		
	1977 CETA	170/546	1650/1783
B. *Non-experimental estimates for displaced workers*			
Bloom (1990)[d]	1984–1985 JTPA, Texas	973	1659
Decker and Corson (1995)	2/88 - 7/88 TAA	−1000	NA
	2/89 - 7/89 TAA	1713	NA
C. *Non-experimental estimates for economically disadvantaged youth participants*			
Cooley et al. (1979)	1969–1971 MDTA	1492	728
Gay and Borus (1980)	1969–1972 Job Corps	−261/180	−1555/−394
Mallar et al. (1982)	1977 Job Corps	2354/2621	NA
Dickinson et al. (1986)	1976 CETA	−1347	449
Bryant and Rupp (1987)	1976 CETA-WE	73(combined)	
Bryant and Rupp (1987)	1976 CETA-WE	1274(combined)	
Bassi et al. (1984)	1977 CETA	−1225/−1614	97/315

[a] Sources: LaLonde (1995, p. 157, Table 1); Barnow (1987, pp. 182–185); Ashenfelter (1978, Tables 4 and 6); Bloom (1990, p. 141, Table 7.6); Kiefer (1979, Table 6.1); Cooley et al. (1979, Table 2); Bassi (1983, Tables 4.3, 4.7, 4.8, and 4.9); Ashenfelter and Card (1985, pp. 658–659); Mallar (1978, Table 1).

[b] MDTA refers to programs funded under the Manpower Development and Training Act, 1962; CETA refers to programs funded under the Comprehensive Employment and Training Act, 1973; JTPA refers to programs funded under the Job Training Partnership Act, 1982; TAA refers to programs funded as part of the Trade Adjustment Assistance Program.

[c] The sets of estimates for each sex refer to the training effect for whites and minorities, respectively.

[d] The Bloom (1990) study was an experimental evaluation. The estimates in the table adjust for non-participation of treatment group members as described in Bloom (1984).

> [al]though these evaluations have all been based on the same datasets, they have produced an extremely wide range of estimated program impacts. In fact, depending on the particular study chosen, one could conclude that CETA programs were quite effective in improving the post-program earnings of participants or, alternatively, that CETA programs reduced the post-program earnings of participants relative to comparable non-participants (Dickinson et al., 1987, pp. 452–453).

Further, different studies of the impact of specific CETA employment and training services exhibit the same variability as the overall program estimates presented in Table 24 (see, e.g., Barnow, 1987, pp. 182–183, Table 3). Five of the six studies summarized in that table also examine the impacts of classroom instruction, on-the-job training, work experience, and public service employment on the earnings of 1976 CETA participants. For example, the estimated effects of OJT for white women in this cohort range from −$295 to $2310 per year. The range of estimates for WE is even larger. Negative training effects are common, but so are large positive impacts.

As discussed in detail in Section 8.4, an important factor contributing to the variability in these non-experimental estimates are differences among analysts' methods of matching. We noted that decisions to match on pre-program earnings at different times substantially affect the estimates. As noted in our discussion on the fallacy of alignment, the problem that arises in these studies is that substantial bias may result when evaluators create comparison groups by matching on serially correlated pre-program outcomes. Matching on such variables alters the properties of the unobservables in the comparison sample in ways that do not guarantee that it will mimic the unobservables of trainees during the post-training period. The bias induced by this practice in the CETA studies can account for their sharply different estimates. Nevertheless, when the estimates from studies most susceptible to this practice are eliminated, the qualitative evidence from the CETA studies is consistent with the experimental evidence from the NJS.

A practical implication of the sensitivity of impact estimates to alternative econometric methods, both experimental and non-experimental, is that cost-benefit analyses of active labor market policies are very fragile. To see the implications of this sensitivity for cost-benefit analyses, consider the following example. Suppose two evaluations of the same program each report that the impacts persist for exactly 8 years. However, the annual impact reported by the first study is $300 per year, while the impact reported by the second study is $700. Assume training costs $2000 per participant. As shown by Table 24, the range of these impacts is consistent with those in the literature. As discussed in Section 2, these costs are typical of government programs. The first evaluation implies that the internal rate of return of the program is 5%, while the second evaluation implies that it is 30%. Readers persuaded by the analysis in the first evaluation would conclude that the program constituted a marginal social investment, whereas those persuaded by the second evaluation would conclude that the program was very productive. This example underscores the importance for policy making of the underlying econometric methodology used

in program evaluations. Modest differences in estimated impacts can have dramatic effects on calculations of the net social benefit of government programs.

Despite the well-documented sensitivity of non-experimental estimates, certain patterns emerge from the non-experimental literature. Government employment and training programs raise the earnings of economically disadvantaged adult women. As shown by Table 24, the estimated impacts are all positive, and many are large relative to the incomes of this population. Further, these impacts are often substantial compared to the costs of these programs which we described in Section 2. Significantly, these results are consistent with the findings in the experimental literature for adult women. In other words, the experimental evaluations, which mostly came after the non-experimental evaluations in time, have led to the same qualitative policy conclusions.

Turning to the impacts for adult males, we observe that they are often smaller and less consistently positive than the impacts for adult women. Accordingly, these estimates suggest that the internal rates of return from these programs are likely lower for males. To illustrate this point consider Ashenfelter's (1978) study of the 1964 MDTA cohort. He reported that CT raised minority males' earnings by $631 and minority females' earnings by $1868. The training cost $8600 (Ashenfelter, 1978, p. 56). If these estimated impacts persisted for the remainder of trainees' working lives, the IRR to training would only be 6% for men, but 22% for women. Because these direct costs include a stipend paid to the trainees, these calculations understate the true IRR. However, they do suggest that these programs constitute a very productive social investment when targeted toward adult women.

As indicated by our discussion of cost-benefit analyses of government programs, these calculations are only suggestive. Many additional considerations besides the earnings impacts affect the IRR of these programs and whether the net benefits are larger when servicing one demographic group compared to another. An important consideration in this regard is the length of the followup period used in the analysis. Ashenfelter's study is relatively unusual in that it followed participants for 5 years after they left the program. By contrast in the CETA studies the followup period usually was less than 2 years. Accordingly, estimates of the IRR from these programs depend crucially on how far into the future analysts project positive shortterm impacts. A second consideration in these IRR calculations is that the foregone earnings cost of participating in training is ignored. These costs are usually larger for adults males. As a result, the gap between the IRR for US programs targeted toward males and females is probably larger than is suggested by the foregoing calculations.

Another factor that may distort simple IRR calculations based on earnings impacts and measures of the average direct cost of training arises because program administrators tend to assign males and females to different services. In the US, males are much more likely to be assigned to receive OJT, which is a less costly service, whereas female participants are more likely to be assigned to receive CT (National Commission for Employment Policy, 1987; Sandell and Rupp, 1988). This practice explains why in Table 20 the internal rates of return estimated for the male participants in the NJS were larger than for females, even

though the earnings impacts shown in Table 22 for the two groups were similar. The males in the NJS were more often assigned to the OJT treatment stream, so the direct costs of servicing them were lower. In the absence of separate measures of the direct costs of these services for males and females, calculations of the IRR of these programs understate the gains from servicing males.

Turning to the non-experimental evaluations of programs for youths, we find that their evidence is also consistent with the results in the experimental literature. The estimated impacts usually are close to zero or even negative. Only one evaluation, that of the US Job Corps program by Mallar et al. (1982), reported substantial positive impacts for youths. However, the earnings impacts during the 4-year followup period are far from sufficient to cover the cost of the program. The modest internal rates of return that have been estimated for this program result from the extrapolation of earnings impacts into the future and from reductions in criminal activity (LaLonde, 1995, p. 164, Table 3). Significantly, these impacts on crime are based on fragile estimates of lower arrest rates for murder (Donohue and Siegelman, 1998). In addition, the comparison group used in this study consisted of non-participants similar to the participants in terms of observable characteristics but drawn from different local labor markets. As explained in Section 8.2, there is now substantial evidence that this approach yields biased estimates of the impact of training. As a result we believe that neither the experimental or non-experimental literatures provide much evidence that employment and training programs improve US youths' labor market prospects.

Over the years both experimental and non-experimental evaluations of government training programs have focused largely on the economically disadvantaged rather than on displaced workers. This focus is in keeping with the emphasis of US employment and training policy on reducing the reliance of low-income persons on various forms of social assistance. Although some of the adult participants in the MDTA and CETA programs would be classified as displaced under the current policy, there have been no separate evaluations of training for displaced workers under these programs that are comparable to those surveyed in Table 24.

As a result, much less is known about the impact of employment and training programs on the earnings of displaced workers. Much of our understanding of how training affects this more advantaged group comes from several demonstrations conducted during the 1980s (Leigh, 1990) as well as from an evaluation of a special program for persons determined to have been displaced by competition from foreign producers (Corson et al., 1993) (see Table 24). Like the MDTA and CETA evaluations, the non-experimental evaluations of these demonstrations find considerable variability in the impact of these training services on different cohorts of displaced workers. But two substantive findings seem clear. First, as is the case for economically disadvantage adults, JSA also is a cost-effective service for displaced workers (Bloom, 1990; Corson et al., 1993). Participants receiving this service have higher earnings because they find jobs sooner than similarly skilled non-participants. Second, participants who have the opportunity to receive CT or OJT derive only modest or no additional benefit from these services.

Given the different objectives of government programs, it also is important to understand how training affects the separate components of earnings, such as employment rates, part-time/full-time status, and hourly wages. A shortcoming of US non-experimental evaluations is that the outcome studied has almost always been annual or quarterly earnings. The CETA and MDTA studies surveyed in Table 24 use annual administrative earnings. These data contain no measures of hours or wages. Further, the employment measure is relatively crude; it reports whether an individual worked in a "covered" job for pay during the year (Card and Sullivan, 1988). Finally, information on the duration of employment or unemployment spells is unavailable. Consequently, by contrast to evaluations of European programs, little is known from non-experimental evaluations of ongoing programs about their impact on employment rates, transition rates out of unemployment or wages. This lack of information makes it difficult to determine whether training raises worker productivity or leads to more stable employment. Much of our knowledge on how US programs affect such outcomes comes from non-experimental evaluations using data from social experiments (see, e.g., Ham and LaLonde, 1996; Eberwein et al., 1997).

10.5. The findings from European evaluations

The European training evaluations are distinct from the US evaluations in several ways. First, they began later in time and only recently has the number become significant. By contrast, the output of such evaluations done by US academics slowed starting in the mid-1980s, although many evaluations continue to be performed by social science consulting firms. This difference in timing results partly from the timing of expanded expenditures on these programs.

Second, European evaluations, particularly those performed outside of the Nordic countries, usually do not use the longitudinal methods commonly used in academic evaluations in the US. Instead, the underlying models are cross-sectional in nature, and control for biases resulting from individual self-selection using parametric methods discussed in Section 7.4. When these evaluations report separate estimates of the impact of training, including and excluding controls for self-selection into training, the estimates controlling for selection usually yield similar or larger estimated impacts than those produced without such controls. As shown by Table 25, this result is seen in evaluations in Austria, Ireland, Norway, Sweden, and the UK. Several authors have noted this finding and have concluded that cross-sectional estimators that fail to account for self-selection into training likely understate the impact of European training programs.

The studies in Sweden and Denmark are generally distinct from other European studies because of their use of longitudinal data and corresponding econometric methods. A factor accounting for this difference is the availability of high quality earnings data from the national "registers." This source of administrative data can yield very large datasets with relatively long panels. For example, the sample used by Westergard-Nielsen (1993) contained more than 30,000 observations covering an 8-year period. This large sample was undoubtedly important in his being able to precisely estimate wage impacts on the

Table 25
Estimated impacts of Canadian and European job training programs (impacts on employment and earnings outcomes)

Country/study author(s)	Outcome studied[a]	Estimator[b]	Program/cohort[c]	Impacts[c]
Austria				
Zweimuller and	Unemployment risk	Probit	ARB-CT: males,1986	0
Winter-Ebmer (1996)		selection		−0.40*
Canada				
Park et al. (1993)	Annual earnings	Diff-in-diff	Canadian jobs strategy	
			CT, 1988	0.09
			CT, 1989	−0.20
			Job Entry, 1988	0.24
			Job Entry, 1989	0.18
			OJT1, 1988	0.06
			OJT1, 1989	−0.11
			OJT2, 1988	0.26*
			OJT2, 1989	−0.01
Denmark				
Jensen et al. (1993)/	Unemployment rate	Panel	AMU: Adults, 1976/1988	0
Westergard-Neilsen (1993)			Poor recent job history	(−)*
	Log hourly wages		Males	0.01*
			Skilled males	0.01
			Unskilled males	0.01*
			Females	0.00
France				
Thierry and Sollogoub (1995)	Employment hazard	MLE	YTP: OJT	(−)*
Bonnal et al. (1997)	Unemployment hazard	MLE	YTP: Males < 26, 1986–1988	
			Without diploma:	
			CT	+*
			WE	0
			OJT	+*

			With certificate:	
			CT	+*
			WE	(−)*
			OJT	+*
	Employment hazard	MLE	Without diploma:	
			CT	(−)*
			WE	0
			OJT	(−)*
			With certificate:	
			CT	(−)*
			WE	+*
			OJT	(−)*
Germany				
Kraus et al. (1997)	Non-employment hazard to stable employment	MLE	AFG: 1992–1994	
			CT-males/females	+*
			OJT-males	0
			OJT-females	+*
Lechner, (1996, 1997)	Unemployment rate	Matching	AFG: 1991–1993	0.01
	Monthly earnings			0.04
Ireland				
Breen (1988, 1991)	Employment rate	Probit	AnCO/FAS: Youth, 1981–1982	
			After exit	0.17*
			After 1 year	0.06
		Selection	After exit	0.25*
			After 1 year	0.04
		Probit	WEP/Teamwork: Youth, 1982–1986	
			After exit	0.23*
			After 1 year	0.26*
		Selection	After exit	0.77*
			After 1 year	0.18
O'Connell and McGinnity (1997)	Employment rate	Probit	Youth < 23, 1992	

Table 25 (*continued*)

Country/study author(s)	Outcome studied[a]	Estimator[b]	Program/cohort[c]	Impacts[c]
			CT	0.16*
			OJT	0.21*
			WE	0.00
	Weekly wage	OLS	CT	0.01*
			OJT	0.02*
			WE	0.00
The Netherlands				
Ridder (1986)	Employment hazard	MLE	E & T Progams: 1979–1981	
			>35 years	0
			<35 years, WE	(−)*
			<35 years, OJT/CT	(−)
	Unemployment hazard	MLE	E & T Progams: 1979–1981	
			>35 years	(−)*
			<35 years, WE	(−)
			<35 years, OJT/CT	(−)
de Koning et al. (1991)	Unemployment hazard	Matching/MLE	CVV-CT:	
			Blue collar trades	+*
			Clerical courses	0
de Koning (1993)	Unemployment rate	OLS	VMA/JOB-OJT:	
			Youth < 25 (JOB)	(−)
			Adults (VMA)	(−)*
Norway				
Torp et al. (1993)	Employment rate	Experiment	CT: All treatments,1991	0.03
		Probit	Training completers only	(−)
		Selection		(−)*
Sweden				
Delander (1978)/	Employment rate	Experiment	ES/Intensified JSA: 1975	
Björklund and Regner (1996)			Job seekers in Eskilstuna	0.13*

	Monthly earnings			0.06
Engstrom et al. (1988)/ Björklund (1993)	Unemployment hazard	MLE	ES: 1983 Displaced workers	0
Björklund (1993, 1994)	Employment rate	OLS	AMS-CT: 1976–1980 16–64 years/unemployed	0.05
		Panel		0.08*
	Log hourly wage	OLS	16–64 years/unemployed	−0.05
		Selection		0.05
		Panel		0.10*
Edin (1988)	Log weekly earnings	Panel	AMS-CT: 1977 Displaced workers	−0.09*
Axelsson (1989)/ Björklund (1993)	Annual earnings (in %)	Panel	AMS-CT: 1981	0.22*
Ackum (1991)	Log hourly wages	OLS	AMS-CT: 1981 Youths < 25	−0.02
		Selection		−0.01
		Panel		−0.05
Andersson (1993)	Annual earnings (in %)	Match/OLS	AMS-CT: 1989–1990 1989 cohort	−0.05*
			1990 cohort	−0.15*
		Match/panel	1989 cohort	−0.02
			1990 cohort	−0.13*
Regner (1996)	Annual earnings (in %)	Match/panel	AMS-CT: 1990 1989 male cohort	0.10
			1990 male cohort	−0.10
			1989 youth cohort	−0.06
			1990 youth cohort	−0.26*
Harkman et al. (1996)	Employment rate	Match/probit	AMS-CT: All, 1993	−0.01
		Match/selection		0.09
	Log hourly wages	Match/OLS		0.02
		Match/Selection		0.05
United Kingdom				
Main and Raffe (1983)	Employment rate	Probit	YOP-Scotland: 1978 Males	0.06
			Females	0.14*

Table 25 (*continued*)

Country/study author(s)	Outcome studied[a]	Estimator[b]	Program/cohort[c]	Impacts[c]
Main (1985)	Employment rate	Probit	YOP-Scotland: 1980	
			Males:	
			All	0.04*
			Disadvantaged[d]	0.03*
			Females:	
			All	0.08*
			Disadvantaged[d]	0.07*
Whitfield and Bourlakis (1991)	Employment rate	Probit	YTS-I:	0.04*
	Log hourly wage	Selection		−0.03
Main and Shelly (1990)	Employment rate	Probit	YTS-I-Scotland: All	0.15*
			Disadvantaged[d]	0.11*
	Log hourly wage	Selection	YTS-I-Scotland:	
			Advantaged[d]	0.20
			Disadvantaged[d]	0.32
Main (1991)	Employment rate	Probit	YTS-I-Scotland:	
			Advantaged[d]	0.14
			Disadvantaged[d]	0.19*
O'Higgins (1994)	Employment rate	Probit	YTS-I: All	0.08*
			Disadvantaged[d]	0.04*
		Selection	YTS-I: All	0.21*
			Disadvantaged[d]	0.09*
			Females	0.28*
Green et al. (1996)	Log hourly wage	Selection	YTS-II:	
			YTS+ certification	0.19
			YTS, no certification	0.02
			YTS, other (e.g., CT)	0.29
Dolton et al. (1992)	Employment rate	Selection	YTS-II: no prior OJT	(−)
			Prior/current OJT	+
	Log hourly wage	Selection	Males age 16 in 1985–1986:	
			WE/OJT	0.05

			CT/OJT	0
			Females age 16 in 1985–1986:	
			WE/OJT	−0.05
			CT/OJT	−0.03
Dolton et al. (1994a)	Unemployment hazard	MLE	YTS-II:	
			To all jobs:	
			Males	(−)*
			Females	(−)*
			To all jobs net of time in YTS:	
			Males	+*
			Females	+*
			To "stable" jobs:	
			Males	0
			Females	+*
Dolton et al. (1994b)	Log hourly wage	Selection	YTS-II: Disadvantaged[d]	
			Males	0.26
			Females	−0.08
White and Lakey (1992)	Employment rate	Experiment	Restart: 1989	0.04*
Payne et al., (1996)	Employment rate	Matching	ET: Adults, 1991–1992	0.22*
	Log hourly wages	Selection		0
	Employment rate	Matching	EA: Adults, 1991–1992	0.04
	Log hourly wages	Selection		0
Dolton and O'Neill (1996a,b)	Unemployment hazard	MLE	Restart: 1989	
			To all jobs	+*
			To stable jobs	0
			To training schemes	+
			To "signing off" UI	+*

[a] Characterizes the class of estimator used in the analysis. "Probit" refers to a univariate probit or logit model that includes one or more indicator variables for training status. "Selection" refers to a parametric econometric model of selection bias that controls for correlation between unobservables in the outcome and participation equations (e.g., Heckman, 1976, 1979; Lee, 1983). "Diff-in-diff" denotes a difference-in-differences estimate that may control for covariates in a regression. "Panel" refers to estimators that exploit longitudinal data, which include difference-in-differences, fixed effects, or autoregressive estimators. "Experiment" refers to the mean difference between outcomes for experimental treatment and control groups. "MLE" refers to standard maximum likelihood methods used for event history data (e.g., Heckman and Singer, 1984; Lancaster, 1990).

[b] Description of employment and training programs: *AMU*: CT usually provide participants with 2–4 weeks of full-time instruction at government training centers. Courses targeted toward both employed and unemployed. Courses partly financed through a payroll tax on employees. *AMS*: CT (Arbetsmarknad-

sutbildning) provided in training centers to unemployed in Sweden. Classroom training programs usually last less than 17 weeks. *AnCO/FAS*: Classroom training lasting less than 6 months. It is sometimes subcontracted to outside or external providers. *ARB*: (Arbeitmarktverwaltung) CT in Austria. *Canadian Jobs Strategy*: consists of several options. CT refers to the feepayers option that allows unemployed adults who receive benefits to enroll at their own expense in approved full-time CT for a period not exceeding 1 year while they continue to receive benefits. Job search requirement waived while in training. Basic skills training not available under "feepayer" option until after 1991. The "Job Entry" option is designed for out-of-school youths and women who have been out of the labor force for three or more years. Separate components for youth and women designed to ease transition into labor force. "OJT1" refers to the "Job Development" option designed for longterm unemployed. "OJT2" refers to "skill shortages" option designed for the unemployed who where "not job ready", who did not meet criteria for other Canadian programs, and who administrators thought would benefit from the program. OJT may last up to 3 years. *CVV*: Vocational Training Centres provide adult participants with vocational CT in blue collar and clerical occupations. *EA*: UK Employment Action program which provides participants with subsidized employment in non-profit or public sector jobs. *East German AFG*: CT programs subsidized under the Work Support Act (Arbeitsforderungsgesetz). *ES*: (Employment Service) services that can include job search assistance (JSA), career counseling, mobility grants, etc. *ET*: UK Employment Training program which provides participants with CT and some OJT opportunities. *Labor Market Training/CT*: Vocational courses provided by Norwegian Directorate of Labour, educational institutions, and private firms. Average duration of such classes was 18 weeks in 1991, although approximately 40% of persons took part in a class lasting 40 or more weeks. *Restart*: April 1987–present, Employment Service provides counseling to all unemployed after 6, 12, and 24 months of unemployment. Assesses claimant's job search behavior and offers advice. May suspend benefits to claimants who are not available for work, who decline offers of assistance, or who fail to attend scheduled interviews. In the Restart experiment, a random sample of unemployed was not required to attend an interview until its 12th month of unemployment. The "treatments" were required to attend the 6 month interview. *YOP*: Youth Opportunities Program 1978–1983. *YTS-I*: Youth Training Scheme 1983–1986 in England and Wales, unless indicated otherwise. *YTS-II*: Youth Training Scheme 1986–1989 in England and Wales, unless indicated otherwise. *VMA/JOB*: provides longterm unemployed adults (VMA) and youth (JOB) with subsidized jobs from private employers. *WEP/Teamwork*: WEP refers to work experience program that provides temporary subsidized jobs with private employer; Teamwork provides the same in a volunteer or community organization. Trainees may be retained by employer when subsidy ends. *YTP*: Youth Training Programs in France.

[c] Impacts measure the percentage effect of the program on earnings or wages and the percentage point impact on employment or unemployment. Both impacts are expressed in decimal form. Results from evaluations of programs on hazard rates out of employment or unemployment are expressed in terms of the effect on the sign of the impact. An asterisk denotes that the impact is statistically significant at the 5% level.

[d] Disadvantaged refers to participants with poor academic qualifications and who reside in local labor markets with high unemployment rates. Advantaged refers to participants with relatively strong academic qualifications (four or more O grades) among non-college bound youth and who reside in local labor markets with low unemployment rates.

order of 1%. The evaluations based on Swedish data usually use a smaller number of observations because they study a random sample of participants and non-participants from the registers, cover a smaller cohort of participants, are limited to a certain geographic section of the country, or discard many non-participants when they create a "matched" comparison group. In Sweden, the ability to use administrative records to match participants to non-participants from the same labor market likely improves the quality of these evaluations. Such matching was impossible in US evaluations that used large administrative datasets (Ashenfelter, 1978, for MDTA; Barnow, 1987, for CETA). Other studies that make use of administrative data include Zweimuller and Winter-Ebmer (1996) for Austria, Ridder (1986) for the Netherlands, and Bonnal et al. (1997) for France. These latter two studies evaluate the effects of training in the context of event history models of labor force dynamics (Flinn and Heckman, 1982).

Evaluations of employment and training programs in the United Kingdom generally use existing general survey data. For example, the evaluations by Whitfield and Bourlakis (1991) and O'Higgins (1994) use the first cohort of the England and Wales Youth Cohort Study (YCS). This survey was administered in three successive years starting in May 1985 to persons who completed their compulsory education during the 1983–1984 academic year.[98] A factor affecting the quality and precision of the estimates in these studies is attrition from the sample. Among those in the first cohort of the YCS only 40% of the original sample responded to all three "sweeps." Similar attrition is reported in existing survey data used in evaluations of the east German programs (Kraus et al., 1997). These experiences underscore the problem of sample attrition when using survey data that does not arise in administrative data obtained from national registers such as those used in the Danish and Swedish studies.

Despite concerns about attrition and the quality of survey responses, an advantage of these survey data is that they contain a much richer set of baseline characteristics on participants and non-participants than is usually available from administrative data sources. For example the UK data contain detailed information on how well the respondent had done in school, including the number of "O" and "A" levels obtained. In addition, it is possible to obtain from both participants and non-participants detailed information on training provided privately by employers. This type of information has generally not been available to evaluators of US programs (but see Gritz, 1993; Heckman and Roselius, 1994). Moreover, these data have enabled evaluators in the UK to look for evidence of heterogeneity in training effects using a wide array of variables that usually has not been available to US evaluators. Finally, these datasets contain local labor market identifiers and as a result several studies have accounted for this variable in their analyses.

A third difference between European and US evaluations is the concentration of these studies on youths. The studies for Austria, Denmark, and Sweden usually include both

[98] The studies by Main and Shelly are based on the comparable Scottish Young Peoples Surveys. The study by Dolton et al. (1992) uses the third cohort of the YCS, which contains individuals who completed their compulsory education during the 1985–1986 academic year.

adults and youths, but nearly all the other studies summarized in Table 25 focus on youth or very young adults. This difference in emphasis reflects policy concerns in Europe about youth unemployment, as compared to policy concerns in the US about the economically disadvantaged of all ages. An advantage of the youth focus of European evaluations is that they provide an opportunity to assess the impact of public sector training interventions on a much less disadvantaged population of youths than is possible in US evaluations. However, surveying the results in the table provides no consistent indication whether these interventions are more or less effective for youth, nor whether more disadvantaged youth benefit more or less from these programs.

A fourth difference between European and US evaluations is that European evaluations place much greater emphasis on measuring the impact of training on hourly wages. As indicated above, this difference reflects the common use of administrative data in US evaluations and the fact that these data almost never contain measures of wages or hours worked. From the perspective of assessing the impact of active labor market policies on human capital accumulation and worker productivity, the European studies potentially shed more light on these questions than is possible in the US studies.

Turning to the estimated impacts presented in the table, we first observe that of the three social experiments conducted in Europe, two tested the impact of employment services along the lines of JSA offered in the United States. Both studies report results that are consistent with those in the US, namely that despite their low costs, access to these services significantly raises employment rates. In the Swedish experiment, unemployed participants received an average of 7.5 h of additional job search assistance compared to 1.5 h received by the control group. Nine months later, the treatments' employment rate was 13 percentage points higher than that of the controls. In the British Restart experiment, a random sample of individuals who had been unemployed for exactly 6 months were assigned to a control group and excused from receiving the 15–25 min interview and counselling session normally required at that time. By contrast, the treatments risked losing their benefits if they failed to attend the interview or demonstrate that they were available for work. Although they could voluntarily request such an interview, the controls were allowed to wait until the next regularly scheduled interview after their twelfth month of unemployment. After 1 year, those assigned to the control group had employment rates that were 4 percentage points lower than those in the treatment group, and for males this impact persisted for at least 5 years (Dolton and O'Neill, 1996b, 1997; Robinson, 1996).[99] The one non-experimental evaluation of JSA was a study of Swedish displaced workers by Engstrom et al. (1988), who found that these services had no significant impact on employment rates.

Among the evaluations summarized in Table 25, we do not observe any pattern that

[99] The original sample contained 8925 persons of which 582 were assigned to a control group. Of the original sample, 5200 persons completed the first 6 month followup survey, of which 323 were controls. Dolton and O'Neill (1996a) found no evidence that this attrition was correlated with a person's experimental status. Dolton and O'Neill matched these survey responses to administrative data (JUVOS) from the Employment Service.

leads us to conclude that any one active labor market policy consistently yields greater employment impacts than any other. Instead, the European evaluations often reveal large and statistically significant effects of any one of these policies on employment rates. This finding is seen directly in the Irish study of Breen (1991), the Swedish study by Björklund (1989), the UK studies by Main and Raffe (1983), Main (1985, 1991), Main and Shelly (1990), and O'Higgins (1994), and indirectly in the Austrian study by Zweimuller and Winter-Ebmer (1996), the French study by Bonnal et al. (1997), and the Dutch study by Ridder (1986). As shown by the table, the estimated employment impacts exceed 10 percentage points in several of these studies.

At the same time, other studies such as the Danish and Norwegian evaluations, the Swedish study by Harkman et al. (1996), and the UK studies by Dolton et al. (1992, 1994b) report much smaller and sometimes even negative impacts of these programs on employment. Although the variability in the impact estimates among studies is reminiscent of the experience with the US CETA evaluations, it is important to observe that these studies are of different cohorts and in some cases of different programs.

Whereas it is common for European evaluations to report that training has significant impact on employment rates, it is relatively uncommon for them to report the same for log wages. In several studies, the point estimates of the impact of training are extremely large, but they are not statistically significant. The largest statistically significant impact reported in the table is by Björklund (1994) who finds that during the late 1970s labor market training in Sweden may have raised hourly wages by 10%. At the same time, he is careful to observe that this finding is sensitive to the econometric method used in the analysis. Moreover, this finding raises the question posed above in Section 10.2 of whether it is plausible that 17 weeks of CT – the standard in Sweden – could result in such a large impact on a trainee's wages. After all, during this period, the impact of a year of formal schooling as measured by a conventional Mincerian wage equation was as low as 2% (Harkman et al., 1996).

In light of this consideration, the other instance of an evaluation reporting a statistically significant impact of training on log wages is more plausible. The Danish study found that 2–4 weeks of vocational classroom training raised the subsequent hourly wages of unskilled male workers by approximately 1%. The point estimate for skilled males was the same, but it was not statistically significant. The point estimates for females were approximately equal to zero, but also not statistically significantly different from 1%. As indicated above, the reason why this study could estimate these impacts so precisely, especially for the males, is because the authors' sample was extremely large.

Although, many of the point estimates of the impact of training on wages are positive, there also are several studies that find either no or negative effects of training on wages. Besides the Danish study referred to above, Whitfield and Bourlakis (1991) and Dolton et al. (1994a) report similar findings for youth in the UK as do Ackum (1991) and Regner (1996) in Sweden. In Sweden several studies also report that training has either no or negative impacts on monthly earnings.

Accordingly, there is little compelling evidence that European active labor market

policies have had a positive impact on participants' wages. By contrast, we have already observed that the case for positive employment effects from these policies is stronger, although there is as yet no consensus on this question. Even if there were a compelling consensus, the question remains whether these employment impacts correspond to an increase in aggregate output or are offset to some extent by displacement of non-participants (Johnson, 1979). Because of the size of these programs as documented in Section 2, because of the emphasis in many European countries on OJT, and because earnings gains from these programs likely are generated through higher employment rates, cost-benefit analyses based on the impact estimates presented in Table 25 probably overstate the net social benefit derived from active labor market policies in Europe.

11. Conclusions

This chapter has examined the effectiveness of active labor market policies and the methods used to evaluate their effectiveness. When these programs are effective they make economically disadvantaged persons less poor, and modestly increase the probability of employment among the unemployed. But the gains from existing programs are not sufficiently large to lift many out of poverty nor to significantly reduce unemployment rates. Further, because these gains, when they occur, appear to arise from increased employment rates instead of wages, they likely overstate the human capital-enhancing benefits of these policies. In Europe, especially, evidence that these programs also result in the displacement of non-participants indicates that the net social benefits of active labor market policies are substantially smaller than are indicated by the impacts from conventional program evaluations.

The evidence we summarize also suggests that it is unlikely that even a substantial increase in government-funded training services will significantly improve the skills in the work force. As indicated above, this finding should not be surprising, because most of these programs cost only a few thousand dollars or less per participant. Although European programs often are more expensive, these costs include stipends paid to participants which do not represent investments in human capital. To expect such programs to raise participants' subsequent annual earnings by several thousand dollars would imply that these social investments consistently have an extraordinary rate of return. A 10% rate of return is high in this literature. Even granting it, a thousand dollars invested in a poor person would only raise annual earnings by $100 per year. A more realistic view of the returns to public-sector-sponsored training would suggest that this type of impact requires an investment that is more than an order of magnitude greater than what is currently being spent on low income and dislocated workers (Heckman et al., 1993).

A major focus of this chapter has been on the methodological lessons learned from 30 years of evaluation activity in the United States and their relevance for the conduct of future evaluations. For brevity, we have left several important issues for discussion elsewhere. In this chapter, we have focused on identifying mean outcomes and in particular the

mean impact of treatment on the treated. Heckman and Smith (1998a) and Heckman et al. (1997c) discuss conditions for recovering distributions of impacts and present evidence on the empirical importance of heterogeneity in impacts in assessing programs. They demonstrate the value of knowing the distribution of program impacts in evaluating the modern welfare state. Heckman et al. (1999) present evidence from the NJS data that persons act on their idiosyncratic response to training, so that the theoretical possibility that we have discussed in this essay is practically important for empirical work in evaluating programs.

We summarize the methodological lessons discussed in this chapter as follows: First, a major development in the field of program evaluation is recognition of the *multiplicity* of the parameters of interest in evaluating employment and training programs. This multiplicity is a consequence of well-documented heterogeneity in the impact of even a single training program. Recognition of this heterogeneity in response among participants and of the possibility that agents participate in programs, at least in part, on the basis of their idiosyncratic responses to them, fundamentally alters intuitions about, and formal properties of, standard econometric estimators. Different parameters require different identifying assumptions, as we demonstrate in our discussion of the conditions for IV to identify "treatment on the treated" rather than LATE in the presence of response heterogeneity. When responses to treatment are heterogeneous, the case for using fixed effect or instrumental variables methods to estimate the parameters commonly sought in evaluation analysis becomes much weaker. Even the case for social experiments has to be qualified significantly if persons enroll in programs at least in part on the basis of their own idiosyncratic response to training.

A second major lesson that flows in part from the first is that the choice of an evaluation method depends on the question being asked in the evaluation and on the economic model generating participation and outcomes. Because both questions and models vary among programs and economic environments, there is no "method of choice" for conducting evaluations. This conclusion is at odds with segments of the current literature which treat matching or, more commonly, fixed effects methods, difference-in-differences or IV as cure-alls for selection problems. Proper choices among alternative experimental and nonexperimental methods should be dictated by the economics of the problem, their relevance to the data in hand, and the evaluation question being addressed. The nature and range of questions being asked by policy makers and researchers make it impossible for a rigorously justified "consensus" to emerge about the proper choice of an estimator to evaluate a social program that is valid in all contexts. All methods for evaluating social programs are based on identifying assumptions that are difficult to test unless additional data about the unobservables in a given study are collected.

There is no universally correct way to construct the counterfactuals needed to evaluate the training programs of the welfare state. Even social experiments are valid only under special assumptions about behavior. We have discussed the interplay between theory, data, and the questions being addressed in an evaluation and how each affects the choice of an estimator. We also have shown that many widely used evaluation strategies – such as choosing comparison groups to make participant and comparison group preprogram earn-

ings histories as alike as possible – only "work" under certain conditions and under other conditions may produce substantially misleading assessments of the program being evaluated.

A third major lesson is that evidence that different estimators produce different estimates, while disappointing, does not necessarily indicate that non-experimental methods fail to measure the appropriate counterfactual. Different estimators solve the selection problem under different assumptions. Only if there is no selection problem and there is no model misspecification problem would all estimators produce the same estimate, up to sampling variation. Robustness studies that show that all methods produce the same estimate only reveal that there is no selection bias.

A fourth major lesson follows from a reexamination of the evidence and issues raised in LaLonde's (1986) paper and Fraker and Maynard's (1987) paper on evaluating non-experimental evaluations. These papers concluded that "...policymakers should be aware that available non-experimental evaluations of employment and training programs may contain large and unknown biases resulting from specification errors" (LaLonde, 1986, p. 617). Nevertheless, some people interpret this work as having proved that conventional econometric program evaluation and model selection procedures are unreliable and cannot be used to produce valid program evaluations. Advocates of social experiments (e.g., Stormsdorfer et al., 1985) and advocates of the robust bounding and sensitivity analyses we briefly survey in Section 7.8 routinely cite it in defense of their methods.

In this chapter, we have reexamined the inferences from this work by drawing on more recent research of Heckman et al. (1996b, 1997a, 1998b). We find that once certain basic principles of data quality are adhered to, selection bias, rigorously defined, is only a small contributor to the bias from using non-experimental data that LaLonde reports in his paper. A far more important bias arises from comparing non-comparable people.

The sources of non-comparability in his study arise from (i) using different surveys or data sources to measure the outcomes and background characteristics of participants and comparison group members; (ii) using participants and comparison group members from different local labor markets; and (iii) using individuals mismatched on personal characteristics. Comparing comparable people goes a long way toward reducing the bias in non-experimental methods reported by LaLonde. This shifts the emphasis in program evaluation away from specifying econometric methods for selection bias and toward more careful construction and weighting of comparison groups. It suggests that in the future, non-experimental comparison groups should be selected to balance the support of the regressors in the comparison group to make it comparable to that in the treatment group. For matching, classical selection bias estimators and non-parametric difference-in-differences estimators, it suggests making the supports of the probability of selection, $P(X)$, coincide in the treatment and comparison groups. This principle should guide both data collection efforts (where stratified sampling of non-participants may be useful) and the analysis of existing datasets.

We also have shown that no econometric or statistical cure-all fixes the problem of fundamentally bad data. Heckman et al. (1998b) demonstrate that econometric selection

estimators and a non-parametric version of difference-in-differences "work" reasonably well for an averaged version of the treatment on the treated parameter when a good comparison group is available. Even the bias from matching is not large. No non-experimental method is particularly effective when a bad comparison group is all that is available. The solution to the evaluation problem lies in both the method *and* the data. The literature on evaluating job training programs has focused largely on methods and not issues of data, taking a passive approach to data collection.

A fifth lesson is that non-experimental evaluations are not necessarily significantly less expensive than experimental evaluations. The low cost of previous non-experimental evaluations resulted from reliance on existing data sources. The importance of high quality data for constructing comparison groups means that credible non-experimental evaluations are likely to be expensive. Existing general survey data and administrative data, which are inexpensively obtained, often contain either too few participants or non-participants, or contain too little information on demographic characteristics or on labor force dynamics. This information has been shown to be important for conducting better non-experimental evaluations and is usually obtained only by collecting costly new survey data. The high cost of previous social experiments results not from administering randomization, but from data collection, careful documentation of the implementation of the program, analysis, and dissemination of reports. These costs are not unique to social experiments, but arise in any careful program evaluation.

A sixth major lesson that emerges from the recent literature is the advantage of using non-parametric econometric methods for program evaluations. The non-parametric approach instructs analysts to compare comparable people. Systematically applied, the non-parametric approach avoids the use of potentially misleading functional forms in constructing counterfactuals.

A seventh major lesson is a better understanding of the benefits and limitations of social experiments. Under ideal conditions, experiments enable us to bypass the need to carefully specify an econometric model or to determine which variables belong in the model. They offer an easily explained procedure for estimating the impact of social programs. In addition, they provide an important benchmark for learning about non-experimental models. Further, even when the ideal conditions are violated, the experimental design enables analysts to obtain a comparison group whose distribution of characteristics is likely similar to those of individuals in the treatment group. Under less than ideal conditions, analysts have to rely on non-experimental methods to estimate parameters of policy interest, but can do so using a better quality comparison group than they could obtain from existing data sources.

Even under ideal conditions, however, the means that can be constructed from a social experiment either by randomizing out people accepted into the program, or randomizing eligibility, identify only a few of the many parameters that can be defined when responses to treatment are heterogeneous, and which are of practical interest to policy makers and social scientists seeking to evaluate active labor market policies. When analysts estimate an evaluation parameter that is not the direct product of the experiment, they must rely on

the same non-experimental methods discussed in Section 7 (Heckman, 1992; Ham and LaLonde, 1996).

The modern case for social experiments usually seeks to recover only one well-defined parameter. This objective is in contrast to the older case that motivated the Negative Income Tax experiments. The older case sought to conduct experiments to recover estimates of the parameters of well-posed economic models that provide the basis for policy analyses of hypothetical programs different from those evaluated by the experiment producing the estimates. Samples generated under the new model for social experiments produce evidence that does not accumulate in the same way as evidence accumulated under the old model, because there is no common basis for comparing the "treatment effects" from one experiment with those from another. Given the nature of the choice-based, endogenously stratified sampling rules used to produce the data used in recent social experiments, it is difficult to use these data to estimate policy-invariant structural parameters that can be used to evaluate a wide variety of programs never previously implemented. Social experiments produced from randomizing out people who applied and were accepted into the program produce knowledge that does not accumulate within the context of economic models unless elaborate non-experimental methods are used to correct for endogenous stratification.

We also have presented evidence on how experiments work in practice. Nearly all social experiments operate in a less than ideal environment and as a result often produce estimates that are not easily interpreted. They are much less effective in evaluating ongoing programs, as illustrated by our discussion of the National JTPA study, than they are in evaluating a new program never previously put in place and for which there are no good substitutes, such as the National Supported Work Demonstration. We draw on the work of Heckman et al. (1998a), who provide evidence that when persons randomized out of the program can find close substitutes for it, the parameter obtained from an experiment differs substantially from the parameter of interest to program evaluators and policy analysts.

An eighth major lesson is that when programs are implemented on a large scale, they may change the prices and opportunities facing everyone in the population. The micro-economic treatment effect literature ignores the effects of programs on the interactions among agents. A convincing evaluation requires embedding the treatment effect framework in a social setting. Drawing on the research of Heckman et al. (1998e) and Davidson and Woodbury (1993), we demonstrate that displacement and general equilibrium effects may be sizeable. The lessons from the treatment effect literature that ignores social interactions can be quite misleading. The challenge in estimating these general equilibrium effects is the challenge of estimating credible general equilibrium models. However, unless the challenge is met, or the social interactions are documented to be unimportant, the output of micro treatment effect evaluations will provide poor guides to public policy.

We conclude this chapter with our recommendations for conducting evaluations based on our best current knowledge. They are: (1) carefully define the parameter of interest; different parameters require different identifying assumptions; (2) compare comparable people; (3) using better data in modeling participation decisions and labor market

outcomes helps a lot. In particular, it is important to measure outcome variables in the same way for participants and non-participants and to draw the treatment and comparison groups from the same local labor markets. In addition, recent evidence suggests that labor force status dynamics represent an important determinant of participation in job training programs; (4) there is no universally "correct" experimental or non-experimental estimator that applies in all contexts. The overwhelming reliance on IV, fixed effects or difference-in-differences and matching estimators in recent research lacks theoretical and empirical justification. In LaLonde's (1986) study, fixed effect estimators produced the most unstable estimates. Evaluators should use economic theory, the available data and prior information to guide the choice of non-experimental estimators, carefully state the conditions under which counterfactual states are generated, and defend their plausibility; (5) expect different estimators to produce different estimates unless there is no selection problem; (6) use experimental methods when possible in evaluating demonstrations of employment and training strategies whose services are not available elsewhere in the community, but collect enough data to test the identifying assumptions that justify experiments. When an experimental design is used to evaluate an ongoing program, analysts should be prepared to use non-experimental methods to answer many important policy questions; (7) the validity of partial equilibrium, microeconomic approaches needs to be confirmed. The estimates from the micro economic treatment effect literature may be very misleading. A more satisfactory approach accounts for the impact of a policy on the interactions of agents in a market economy.

References

Aakvik, A. (1998), "Estimating the employment effects of education for disabled workers in Norway", Unpublished manuscript (University of Chicago).

Ackum, S. (1991), "Youth unemployment, labour market programs and subsequent earnings", Scandinavian Journal of Economics 93 (4): 531–543.

Amemiya, T. (1985), Advanced econometrics (Harvard University Press, Cambridge, MA).

Anderson, K., R. Burkhauser, J. Raymond and C. Russell (1991), "Mixed signals in the Job Training Partnership Act", Growth and Change 22 (3): 32–48.

Anderson, K., R. Burkhauser and J. Raymond (1993), "The effect of creaming on placement rates under the Job Training Partnership Act", Industrial and Labor Relations Review 46 (4): 613–624.

Andersson, H. (1993), "Choosing among alternative nonexperimental methods for estimating the impact of training: new Swedish evidence", Unpublished manuscript (Swedish Institute for Social Research, Stockholm University).

Andrews, D. and M. Schafgans (1998), "Semiparametric estimation of a sample selection model", Review of Economic Studies 56 (3).

Ashenfelter, O. (1978), "Estimating the effect of training programs on earnings", Review of Economics and Statistics 6 (1): 47–57.

Ashenfelter, O. (1979), "Estimating the effect of training programs on earnings with longitudinal data", in: F. Bloch, ed., Evaluating manpower training programs (JAI Press, Greenwich, CT) pp. 97–117.

Ashenfelter, O. and D. Card (1985), "Using the longitudinal structure of earnings to estimate the effect of training programs", Review of Economics and Statistics 67 (3): 648–660.

Ashenfelter, O. and C. Rouse (1995), "Schooling, intelligence and income in America: cracks in the bell curve", Unpublished manuscript (Princeton University).

Axelsson, R. (1989), "Svensk arbetsmarknadsutbildning - en kvantitativ analys av dess effekter", Umeå economic studies (Umeå University).

Balestra, P. and M. Nerlove (1966), "Pooling cross section and time series data in the estimation of a dynamic model: the demand for natural gas", Econometrica 34 (1): 585–612.

Balke, A. (1995), "Probabilistic counterfactuals: semantics, computation, and applications", Technical report R-242 (UCLA).

Balke, A. and J. Pearl (1993), "Nonparametric bounds on causal effects from partial compliance data", Technical report R-199 (UCLA).

Balke, A. and J. Pearl (1997), "Bounds on treatment effects from studies with imperfect compliance", Journal of the American Statistical Association 92 (439): 1171–1176.

Baltagi, B. (1995), Econometric analysis of panel data (Wiley, New York).

Barnow, B. (1987), "The impact of CETA programs on earnings: a review of the literature", Journal of Human Resources 22: 157–193.

Barnow, B., G. Cain and A. Goldberger (1980), "Issues in the analysis of selectivity bias", in: E. Stromsdorfer and G. Farkas, eds., Evaluation studies, Vol. 5 (Sage Publications, Beverly Hills, CA) pp. 290–317.

Barron, J., D. Black and M. Lowenstein (1989), "Job matching and on-the-job training", Journal of Labor Economics 7 (1): 1–19.

Barron, J., M. Berger and D. Black (1997), "How well do we measure training?" Journal of Labor Economics 15 (3): 507–528.

Bassi, L. (1983), "The effect of CETA on the post-program earnings of participants", Journal of Human Resources 18 (Fall): 539–556.

Bassi, L. (1984), "Estimating the effect of training programs with non-random selection", Review of Economics and Statistics 66 (1): 36–43.

Bassi, L. and O. Ashenfelter (1986), "The effects of direct job creation and training programs on low-skilled workers", in: S. Danziger and D. Weinberg, eds., Fighting poverty (Harvard University Press, Cambridge, MA) pp. 133–151.

Bassi, L., M. Simms, L. Burnbridge and C. Betsey (1984), "Measuring the effect of CETA on youth and the economically disadvantaged", Final report prepared for the US Department of Labor under contract no. 20-11-82-19 (The Urban Institute, Washington, DC).

Begg, I., A. Blake and B. Deakin (1991), "YTS and the labour market", British Journal of Industrial Relations 29 (2): 223–236.

Bell, S. and C. Reesman (1987), AFDC Homemaker-Home Health Aide demonstrations: trainee potential and performance (Abt Associates, Cambridge, MA).

Bell, S., L. Orr, J. Blomquist and G. Cain (1995), Program applicants as a comparison group in evaluating training programs (W.E. Upjohn Institute for Employment Research, Kalamazoo, MI).

Bera, A., C. Jarque and L. Lee (1984), "Testing the normality assumption in limited dependent variable models", International Economic Review 25 (3): 563–578.

Berry, D. and B. Fristedt (1985), Bandit problems (Chapman and Hall, London).

Björklund, A. (1989), "Evaluations of training programs: experiences and suggestions for future research", Discussion paper no. 89 (Wissenschaftszentrum, Berlin).

Björklund, A. (1993), "The Swedish experience", in: K. Jensen and P.K. Madsen, eds., Measuring labour market measures (Ministry of Labour, Copenhagen, Denmark) p. 243–263.

Björklund, A. (1994), "Evaluations of Swedish labor market policy", International Journal of Manpower 15 (5, part 2): 16–31.

Björklund, A. and R. Moffitt (1987), "Estimation of wage gains and welfare gains in self-selection models", Review of Economics and Statistics 69 (1): 42–49.

Björklund, A. and H. Regner (1996), "Experimental evaluation of European labour market policy", in: G.

Schmid, J. O'Reilly and K. Schömann, eds., International handbook of labour market policy and evaluation (Edward Elgar, Aldershot, UK) pp. 89–114.

Bloom, H. (1984), "Accounting for no-shows in experimental evaluation designs", Evaluation Review 82 (2): 225–246.

Bloom, H. (1990), Back to work: testing reemployment services for displaced workers (W.E. Upjohn Institute for Employment Research, Kalamazoo, MI).

Bloom, H. and M. McLaughlin (1982), CETA training programs: do they work for adults? Joint report (CBO-NCEP, Washington, DC).

Bloom, H., L. Orr, G. Cave, S. Bell and F. Doolittle (1993), The National JTPA Study: title II-A impacts on employment and earnings at 18 months (Abt Associates, Bethesda, MD).

Bonnal, L., D. Fougere and A. Serandon (1997), "Evaluating the impact of French employment policies on individual labour market histories", Review of Economic Studies 64 (4): 683–713.

Bound, J., C. Brown, G. Duncan and W. Rodgers (1994), "Evidence on the validity of cross-sectional and longitudinal labor market data", Journal of Labor Economics 12 (3): 345–368.

Bradley, S. (1994), "The Youth Training Scheme: A critical review of the evaluation literature", International Journal of Manpower 16 (4): 30–56.

Breen, R. (1988), "The work experience program in Ireland", International Labour Review 127 (4): 429–444.

Breen, R. (1991), "Assessing the effectiveness of training and temporary employment schemes: some results from the youth labour market", The Economic and Social Review 22 (3): 177–198.

Brown, R. (1979), Assessing the effects of interview nonresponse on estimates of the impact of supported work (Mathematica Policy Research, Princeton, NJ).

Browning, E. (1987), "On the marginal welfare cost of taxation", American Economic Review 77 (1): 11–23.

Bryant, E. and K. Rupp (1987), "Evaluating the impact of CETA on participant earnings", Evaluation Review 11: 473–492.

Burtless, G. (1995), "The case for randomized field trials in economic and policy research", Journal of Economic Perspectives 9 (2): 63–84.

Butler, R. and J. Heckman (1977), "Government's impact on the labor market status of black Americans: a critical review", in: Equal rights and industrial relations (Industrial Relations Research Association, Madison, WI) pp. 235–281.

Cain, G. (1975), "Regression and selection models to improve nonexperimental comparisons", in: C. Bennett and A. Lumsdaine, eds., Evaluation and experiment (Academic Press, New York) pp. 297–317.

Calmfors, L. (1994), "Active labor market policy and unemployment - a framework for the analysis of crucial design features", OECD Economic Studies 22 (1): 7–47.

Cameron, S. and J. Heckman (1998), "Life cycle schooling and dynamic selection bias: models and evidence for five cohorts of American males", Journal of Political Economy 106 (2): 262–333.

Campbell, D. and J. Stanley (1963), "Experimental and quasi-experimental designs for research on teaching", in: N. Gage, ed., Handbook of research on teaching (Rand McNally, Chicago, IL) pp. 171–246.

Campbell, D. and J. Stanley (1966), Experimental and quasi-experimental designs for research (Rand McNally, Chicago, IL).

Card, D. (1995), "Earnings, schooling and ability revisited", Working paper no. 4832 (NBER, Cambridge, MA).

Card D. and D. Sullivan (1988), "Measuring the effects of CETA participation on movements in and out of employment", Econometrica 56 (3): 497–530.

Cave, G., H. Bos, F. Doolittle and C. Toussaint (1993), JOBSTART: final report on a program for school dropouts (Manpower Demonstration Research Corporation, New York).

Chamberlain, G. (1984), "Panel data", in: Z. Griliches and M. Intriligator, eds, Handbook of econometrics (North-Holland, Amsterdam) pp. 1248–1318.

Chickering, D. and J. Pearl (1996), "A clinician's tool for analyzing non-compliance", in: Proceedings of the National Conference on Artificial Intelligence (AAAI-96) (Morgan Kauffman, Boston, MA) pp. 1269–1276.

Chipman, J. and J. Moore (1976), "Why an increase in GNP need not imply an improvement in potential welfare", Kyklos 29: 391–418.

Cochran, W. and D. Rubin (1973), "Controlling bias in observational studies", Sankyha 35: 417–446.
Cooley, T., T. McGuire and E. Prescott (1979), "Earnings and employment dynamics of manpower trainees: an exploratory econometric analysis", in: R. Ehrenberg, ed., Research in labor economics, Vol. 4, Suppl. 2 (JAI Press, Greenwich, CT) pp. 119–147.
Cook, T. and D. Campbell (1979), Quasi-experimentation: design and analysis issues for field settings (Houghton-Mifflin, Boston, MA).
Corson, W., P. Decker, P. Gleason and W. Nicholson (1993), International trade and worker dislocation: evaluation of the trade adjustment assistance program (Mathematica Policy Research, Princeton, NJ).
Cosslett, S. (1983), "Distribution-free maximum likelihood estimator of the binary choice model", Econometrica 51 (3): 765–782.
Couch, K. (1992), "New evidence on the long-term effects of employment and training programs", Journal of Labor Economics 10 (4): 380–388.
Davidson, C. and S. Woodbury (1993), "The displacement effect of reemployment bonus programs", Journal of Labor Economics 11 (4): 575–605.
Davidson, C. and S. Woodbury (1995), "Wage subsidies for dislocated workers", Unpublished manuscript (W.E. Upjohn Institute for Employment Research, Kalamazoo, MI).
Decker, P. and W. Corson (1995), "International trade and worker displacement: evaluation of the trade adjustment assistance program", Industrial and Labor Relations Review 48 (4): 758–774.
de Koning, J. (1993), "Measuring the placement effects of two wage subsidy schemes for the long term unemployed", Empirical Economics 18: 447–468.
de Koning, J., M. Koss and A. Verkaik (1991), "A quasi-experimental evaluation of the vocational training centre for adults", Environment and Planning C: Government and Policy 9: 143–153.
Delander, L. (1978), "Studier kring den arbetsformedlande verksamheten" (Studies of the Swedish Employment Office) in SOU: 60.
Devine, T. and J. Heckman (1996), "The structure and consequences of eligibility rules for a social program", in: S. Polachek, ed., Research in labor economics, Vol. 15 (JAI Press, Greenwich, CT) pp. 111–170.
Dickinson, K., T. Johnson and R. West (1984), "An analysis of the impact of CETA programs on participants' earnings", Report prepared for the US Department of Labor under contract no. 20-06-82-21 (SRI International, Menlo Park, CA).
Dickinson, K., T. Johnson and R. West (1986), "An analysis of the impact of CETA on participants' earnings", Journal of Human Resources, 21: 64–91.
Dickinson, K., T. Johnson and R. West (1987), "An analysis of the sensitivity of quasi-experimental net estimates of CETA programs", Evaluation Review 11: 452–472.
Dolton, P. (1993), "The economics of youth training in Britain", Economic Journal 103 (420): 1261–1278.
Dolton, P. and D. O'Neill (1996a), "Unemployment duration and the Restart effect: some experimental evidence", Economic Journal 106 (435): 387–400.
Dolton, P. and D. O'Neill (1996b), "The Restart effect and the return to full-time stable employment", Journal of the Royal Statistical Society Series A 159 (2): 275–288.
Dolton, P. and D. O'Neill (1997), "The long-run effect of unemployment monitoring and work search programs: some experimental evidence from the U.K.", Unpublished monograph (University of Newcastle-upon-Tyne).
Dolton, P., G. Makepeace and J. Treble (1992), "Public- and private-sector training of young people in Britian", in: L. Lynch, ed., Training and the private sector (University of Chicago Press, Chicago, IL) pp. 261–281.
Dolton, P., G. Makepeace and J. Treble (1994a), "The Youth Training Scheme and the school-to-work transition", Oxford Economic Papers 46 (4): 629–657.
Dolton, P., G. Makepeace and J. Treble (1994b), "The wage effect of YTS: evidence from YCS", Scottish Journal of Political Economy 41 (4): 444–453.
Donohue, J. and P. Siegelman (1998), "Allocating resources among prisons and social programs in the battle against crime", Journal of Legal Studies 27 (1): 1–43.
Doolittle, F. and L. Traeger (1990), Implementing the National JTPA Study (Manpower Demonstration Research Corporation, New York).

Eberwein, C., J. Ham and R. LaLonde (1997), "The impact of classroom training on the employment histories of disadvantaged women: evidence from experimental data", Review of Economic Studies 64 (4): 655–682.

Edin, P.-A. (1988), "Individual consequences of plant closures", PhD dissertation (Uppsala University).

Engstrom, L., K. Lofgren and O. Westerlund (1988), "Intensified employment services, unemployment duration and unemployment risks", Economic studies no. 186 (Umeå University).

Farber, H. and R. Gibbons (1994), "Learning and wage dynamics", Unpublished manuscript (Princeton University).

Fay, R. (1996), "Enhancing the effectiveness of active labour market policies: evidence from programme evaluations in OECD Countries", Occasional papers no. 18 (Labour market and social policy, OECD, Paris).

Fechner, G. (1860), Elemente der psychophysik (Breitkopf and Härtel, Leipzig, Germany).

Finifter, D. (1987), "An approach to estimating the net earnings impact of federally subsidized employment and training programs", Evaluation Review 11 (4): 528–547.

Fisher, R. (1935), Design of experiments (Hafner, New York).

Flinn, C. and J. Heckman (1982), "New methods for analyzing structural models of labor force dynamics", Journal of Econometrics 18 (1): 115–168.

Forslund, A. and A. Krueger (1997), "An evaluation of the Swedish active labor market policy: new and received wisdom", in: R. Freeman, R. Topel and B. Swedenburg, eds, The welfare state in transition (The University of Chicago Press for NBER, Chicago, IL).

Fraker, T. and R. Maynard (1987), "The adequacy of comparison group designs for evaluations of employment-related programs", Journal of Human Resources 22 (2): 194–227.

Friedlander, D. and G. Hamilton (1993), The saturation work initiative model in San Diego: a five-year follow-up study (Manpower Demonstration Research Corporation, New York).

Friedlander, D and P. Robbins (1995), "Evaluating program evaluations: new evidence on commonly used nonexperimental methods", American Economic Review 85 (4): 923–937.

Friedlander, D., G. Hoertz, J. Quint and J. Riccio (1985), Arkansas, the demonstration of state work/welfare initiatives: the final report on the WORK program in two counties (Manpower Demonstration Research Corporation, New York).

Friedlander, D., D. Greenberg and P. Robins (1997), "Evaluating government training programs for the economically disadvantaged", Journal of Economic Literature 35 (4): 1809–1855.

Gay, R. and M. Borus (1980), "Validating performance indicators for employment and training programs", Journal of Human Resources 15: 29–48.

Geraci, V. (1984), "Short-term indicators of job training program effects on long-term participant earnings, Report prepared for US Department of Labor under contract no. 20-48-82-16.

Glynn, R. and D. Rubin (1986), "Selection modeling versus mixture modeling", in: H. Wainer, ed., Drawing inferences from selected samples (Springer-Verlag, Berlin).

Goldberger, A. (1972), "Selection bias in evaluating treatment effects", Discussion paper no. 123-172 (Institute for Research on Poverty, University of Wisconsin).

Goldman, B., D. Friedlander and D. Long (1986), California, the demonstration of state work/welfare initiatives: final report on the San Diego job search and work experience demonstration (Manpower Demonstration Research Corporation, New York).

Gramlich, E. and B. C. Ysander (1981), "Relief work and grant displacement in Sweden", in: G. Eliasson, B. Holmlund and F. Stafford, eds., Studies in labor market behavior: Sweden and the United States (The Industrial Research Institute, Stockholm, Sweden) pp. 139–166.

Green, F., M. Hoskins and S. Montgomery (1996) "The effects of company training, further education and the youth training scheme on the earnings of young employees", Oxford Bulletin of Economics and Statistics 58 (3) 469–488.

Greenberg, D. (1997), "The leisure bias in cost–benefit analyses of employment and training programs", Journal of Human Resources 32 (2): 413–439.

Greenberg, D. and M. Wiseman (1992), "What did the OBRA demonstrations do?" in: C. Manski and I.

Garfinkel, eds., Evaluating welfare and training programs (Harvard University Press, Cambridge, MA) pp. 25–75.
Gritz, M. (1993), "The impact of training on the frequency and duration of employment", Journal of Econometrics 57 (1–3): 21–51.
Grossman, J., R Maynard and J. Roberts (1985), Reanalysis of the effects of selected employment and training programs for welfare recipients (Mathematica Policy Research, Princeton, NJ).
Gueron, J. (1990), "Work and welfare: lessons on employment programs", Journal of Economic Perspectives 4 (1): 79–98.
Hahn, J., P. Todd and W. van der Klaauw (1998), "Estimation of treatment effects with a quasi-experimental regression-discontinuity design with application to evaluating the effect of federal antidiscrimination laws on minority employment in small U.S. firms", Unpublished manuscript (University of Pennsylvania).
Ham J. and R. LaLonde (1990), "Using social experiments to estimate the effect of training on transition rates", in: J. Hartog, G. Ridder and J. Theeuwes, eds., Panel data and labor market studies (North-Holland, Amsterdam) pp. 157–172.
Ham J. and R. LaLonde (1996), "The effect of sample selection and initial conditions in duration models: evidence from experimental data", Econometrica 64 (1): 175–205.
Hamermesh, D. (1971), Economic aspects of manpower training programs (Lexington Books, Lexington, MA).
Hamermesh, D. (1993), Labor demand (Princeton University Press, Princeton, NJ).
Harberger, A. (1971), "Three basic postulates for applied welfare economics", Journal of Economic Literature 9 (3): 785–797.
Harkman, A., F. Jansson and A. Tamas (1996), "Effects, defects, and prospects – an evaluation of labor market training in Sweden", Unpublished manuscript (Research Unit, Swedish National Labour Market Board).
Haveman, R. and D. Saks (1985), "Transatlantic lessons for employment and training policy", Industrial Relations 24 (2): 20–36.
Heckman, J. (1976), "Simultaneous equations models with continuous and discrete endogenous variables and structural shifts", in: S. Goldfeld and R. Quandt, eds., Studies in nonlinear estimation (Ballinger, Cambridge, MA).
Heckman, J. (1978), "Dummy endogenous variables in a simultaneous equations system", Econometrica 46 (4): 931–959.
Heckman, J. (1979), "Sample selection bias as a specification error", Econometrica 47 (1): 153–161.
Heckman, J. (1980), "Addendum to sample selection bias as a specification error", in: E. Stromsdorfer and G. Farkas, eds., Evaluation studies review annual, Vol. 5 (Sage, San Francisco, CA) pp. 970–995.
Heckman, J. (1990), "Varieties of selection bias", American Economic Review 80 (2): 313–318.
Heckman, J. (1992), "Randomization and social policy evaluation", in: C. Manski and I. Garfinkel, eds., Evaluating welfare and training programs (Harvard University Press, Cambridge, MA) pp. 201–230.
Heckman, J. (1996), "Randomization as an instrumental variable", Review of Economics and Statistics 78 (2): 336–341.
Heckman, J. (1997), "Instrumental variables: a study of implicit behavioral assumptions in one widely used estimator", Journal of Human Resources, 32 (3): 441–461.
Heckman, J. (1998a), "The economic evaluation of social programs", in: J. Heckman and E. Leamer, eds., Handbook of econometrics, Vol. 5 (Elsevier, Amsterdam), in press.
Heckman, J., ed. (1998b), Performance standards in a government bureaucracy: analytical essays on the JTPA performance standards system (W.E. Upjohn Institute for Employment Research, Kalamazoo, MI).
Heckman, J. (1998c), "A unified matching and weighting framework for all evaluation estimators", Unpublished manuscript (University of Chicago).
Heckman, J. and G. Borjas (1980), "Does unemployment cause future unemployment? Definitions, questions and answers from a continuous time model of heterogeneity and state dependence", Economica 47 (187): 247–283.
Heckman, J. and B. Honoré (1990), "The empirical content of the Roy model", Econometrica 58 (5): 1121–1149.
Heckman, J. and J. Hotz (1989), "Choosing among alternative methods of estimating the impact of social

programs: the case of manpower training", Journal of the American Statistical Association 84 (408): 862–874.

Heckman, J. and T. MaCurdy (1986), "Labor econometrics", in: Z. Griliches and M. Intriligator, eds., Handbook of econometrics (North-Holland, Amsterdam) pp. 1917–1977.

Heckman, J. and R. Robb (1982), "The longitudinal analysis of earnings", Unpublished manuscript (University of Chicago).

Heckman, J. and R. Robb (1985a), "Alternative methods for evaluating the impact of interventions", in: J. Heckman and B. Singer, eds., Longitudinal analysis of labor market data (Cambridge University Press for Econometric Society Monograph Series, New York) pp. 156–246.

Heckman, J. and R. Robb (1985b), "Alternative methods for evaluating the impact of interventions: an overview", Journal of Econometrics 30 (1,2): 239–267.

Heckman, J. and R. Robb (1986a), "Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes", in: H. Wainer, ed., Drawing inferences from selected samples (Springer-Verlag, Berlin) pp. 63–107.

Heckman, J. and R. Robb (1986b), "Alternative identifying assumptions in econometric models of selection bias", in: G. Rhodes, ed. Advances in econometrics. Vol. 5 (JAI Press, Greenwich, CT) pp. 243–287.

Heckman, J. and R. Roselius (1994), "Evaluating the impact of training on the earnings and labor force status of young women: better data help a lot", Unpublished manuscript (University of Chicago).

Heckman, J. and G. Sedlacek (1985), "Heterogeneity, aggregation and market wage functions: an empirical model of self-selection in the labor market", Journal of Political Economy 98 (6): 1077–1125.

Heckman, J. and B. Singer (1984), "A method for minimizing the impact of distributional assumptions in econometric models for duration data", Econometrica 52 (2): 271–320.

Heckman, J. and J. Smith (1993), "Assessing the case for randomized evaluation of social programs", in: K. Jensen and P.K. Madsen, eds., Measuring labour market measures (Ministry of Labour, Copenhagen, Denmark) pp. 35–96.

Heckman, J. and J. Smith (1995), "Assessing the case for social experiments", Journal of Economic Perspectives 9 (2): 85–100.

Heckman, J. and J. Smith (1998a), "Evaluating the welfare state", in: S. Strom, ed., Econometrics and economics in the 20th century: the Ragnar Frisch centenary (Cambridge University Press for Econometric Society Monograph Series, New York).

Heckman, J. and J. Smith (1998b), "The sensitivity of experimental impact estimates: evidence from the National JTPA Study", in: R. Freeman and L. Katz, eds., Youth employment and unemployment in the OECD countries (University of Chicago Press for NBER, Chicago, IL) in press.

Heckman, J. and J. Smith (1998c), "The performance of performance standards: the effects of JTPA performance standards on efficiency, equity and participant outcomes", Unpublished manuscript (University of Chicago).

Heckman, J. and J. Smith (1998d), "The determinants of participation in a social program: evidence from the job training partnership act", Unpublished manuscript (University of Chicago).

Heckman, J. and J. Smith (1998e), "The sensitivity of nonexperimental evaluation estimators: a simulation study", Unpublished manuscript (University of Chicago).

Heckman, J. and J. Smith (1999), "The pre-program dip and the determinants of program participation in a social program: implications for simple program evaluation strategies", Economic Journal, in press.

Heckman, J. and P. Todd (1994), "Interpreting standard measures of selection bias", Unpublished manuscript (University of Chicago).

Heckman, J. and E. Vytlacil (1999a), "Local instrumental variables and latent variable models for identifying and bounding treatment effects", Proceedings of the National Academy of Sciences USA 96: 4730–4734.

Heckman, J. and E. Vytlacil (1999b), "The relationship between treatment parameters within a latent variable framework", Economics Letters, in press.

Heckman, J. and K. Wolpin (1976), "Does the contract compliance program work? An analysis of Chicago data", Industrial and Labor Relations Review 19: 415–433.

Heckman, J., R. Roselius and J. Smith (1993), "U.S. education and training policy: a re-evaluation of the

underlying assumptions behind the 'new consensus'", in: A. Levenson and L. Solomon, eds., Labor markets, employment policy and job creation (Milken Institute for Job and Capital Formation, Santa Monica, CA) pp. 83–121.

Heckman, J., M. Khoo, R. Roselius and J. Smith (1996a), "The empirical importance of randomization bias in social experiments: evidence from the national JTPA study", Unpublished manuscript (University of Chicago).

Heckman, J., H. Ichimura, J. Smith and P. Todd (1996b), "Sources of selection bias in evaluating social programs: an interpretation of conventional measures and evidence on the effectiveness of matching as a program evaluation method", Proceedings of the National Academy of Sciences USA 93 (23): 13416–13420.

Heckman, J., J. Smith and C. Taber (1996c), "What do bureaucrats do? The effects of performance standards and bureaucratic preferences on acceptance into the JTPA program", in: G. Libecap, ed., Reinventing government and the problem of bureaucracy, Vol. 6, Advances in the study of entrepreneurship, innovation and economic growth (JAI Press, Greenwich, CT) pp. 191–218.

Heckman, J., H. Ichimura and P. Todd (1997a), "Matching as an econometric evaluation estimator: evidence from evaluating a job training program", Review of Economic Studies 64 (4): 605–654.

Heckman, J., L. Lochner, J. Smith and C. Taber (1997b), "The effects of government policies on human capital investment and wage inequality", Chicago Policy Review 1 (2): 1–40.

Heckman, J., J. Smith and N. Clements (1997c), "Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts", Review of Economic Studies 64 (4): 487–535.

Heckman, J., N. Hohmann and J. Smith with M. Khoo (1998a), "Substitution and dropout bias in social experiments: evidence from an influential social experiment", Quarterly Journal of Economics, in press.

Heckman, J., H. Ichimura, J. Smith and P. Todd (1998b), "Characterizing selection bias using experimental data", Econometrica 66: 1017–1098.

Heckman, J., H. Ichimura and P. Todd (1998c), "Matching as an econometric evaluation estimator", Review of Economic Studies 65 (2): 261–294.

Heckman, J., L. Lochner and C. Taber (1998d), "Explaining rising wage inequality: explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents", Review of Economic Dynamics 1 (1): 1–64.

Heckman, J., L. Lochner and C. Taber (1998e), "General equilibrium treatment effects: a study of tuition policy", American Economic Review 88 (2): 381–386.

Heckman, J., J. Smith and C. Taber (1998f), "Accounting for dropouts in evaluations of social programs", Review of Economics and Statistics 80 (1): 1–14.

Heckman, J., J. Smith and P. Todd (1999), "The evaluation problem", Unpublished manuscript (University of Chicago).

Heinrich, C. (1998), "The role of performance standards in JTPA program administration and service delivery at the local level", in: J. Heckman, ed., Performance standards in a government bureaucracy: analytical essays on the JTPA performance standards system (W.E. Upjohn Institute for Employment Studies, Kalamazoo, MI) in press.

Holland, P. (1986), "Statistics and causal inference", Journal of the American Statistical Association 81 (396): 945–960.

Holland, P. (1988), "Causal inference, path analysis and recursive structural equation models", in: C. Clogg, ed., Sociological methodology (American Sociological Association, Washington, DC) pp. 449–484.

Hollister, R. and D. Freedman (1988), "Special employment programmes in OECD countries", International Labour Review 127 (3): 317–334.

Hollister, R., P. Kemper and R. Maynard (1984), The National Supported Work demonstration (University of Wisconsin Press, Madison, WI).

Honoré, B. and E. Kyriazidou (1998), "Panel data discrete choice models with lagged dependent variables", Unpublished manuscript (University of Chicago).

Hotz, V.J. (1992), "Designing an evaluation of the Job Training Partnership Act", in: C. Manski and I. Garfinkel, eds., Evaluating welfare and training programs (Harvard University Press, Cambridge, MA) pp. 76–114.

Hsiao, C. (1986), Analysis of panel data (Cambridge University Press for Econometric Society Monograph Series, Cambridge, UK).

Hutchinson, G. and A. Church (1989), "Wages, unions, the Youth Training Scheme and the Young Workers Scheme", Scottish Journal of Political Economy 36 (2): 160–182.

Ichimura, H. (1993), "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models", Journal of Econometrics 58 (1,2): 71–120.

Imbens, G. and J. Angrist (1994), "Identification and estimation of local average treatment effects", Econometrica 62 (4): 467–476.

Imbens, G. and T. Lancaster (1996), "Case-control studies with contaminated controls", Journal of Econometrics 71 (1,2): 145–160.

Jensen, P., P. Pederson, N. Smith and N. Westergard-Nielson (1993), "The effects of labor market training on wages and unemployment: some Danish results", in: H. Bunzel, P. Jensen and N. Westergard-Nielson, eds., Panel data and labour market dynamics, contributions to economic analysis no. 222 (North Holland, Amsterdam) pp. 311–331.

Johnson, G. (1979), "The labor market displacement effect in the analysis of the net impact of manpower training programs", in: F. Bloch, ed., Evaluating manpower training programs, Suppl. 1 (JAI Press, Greenwich, CT) pp. 227–254.

Johnson, G. and R. Layard (1986), "The natural rate of unemployment: explanation and policy", in O. Ashenfelter and R. Layard, eds., Handbook of labor economics, Vol. 2 (North-Holland, Amsterdam) pp. 921–999.

Johnson, G. and J. Tomola (1977), "The fiscal substitution effects of alternative approaches to public service employment", Journal of Human Resources 12 (1): 3–26.

Kane, T. (1994), "College entry by blacks since 1970: the role of college costs, family background and the return to education", Journal of Political Economy 102 (5): 878–912.

Kane, T. and C. Rouse (1993), "Labor market returns to two- and four-year college", American Economic Review 85 (3): 600–614.

Kemper, P., D. Long and C. Thornton (1981), The supported work evaluation: final cost benefit analysis (Manpower Demonstration Research Corporation, New York).

Kemper, P., D. Long and C. Thornton (1984), "A benefit–cost analysis of the supported work experiment", in: R. Hollister, P. Kemper and R. Maynard, eds., The National Supported Work demonstration (University of Wisconsin Press, Madison, WI) pp. 239–285.

Kemple, J., D. Friedlander and V. Fellerath (1995), Florida's project independence: benefits, costs and two-year impacts of Florida's JOBS program (Manpower Demonstration Research Corporation, New York).

Kiefer, N. (1979), The economic benefits of four employment and training programs (Garland Publishing, New York).

Knox, V., P. Auspos, J. Hunter-Manns, C. Miller and A. Orenstein (1997), Making welfare work: 18-month impacts of Minnesota's family investment program (Manpower Demonstration Research Corporation, New York).

Kornfeld, R. and H. Bloom (1996), "Measuring the impacts of social programs on the earnings and employment of low income persons: do UI wage records and surveys agree?" Unpublished manuscript (Abt Associates, Bethesda, MD).

Kraus, F., P. Puhani and V. Steiner (1997), "Employment effects of publically financed training programs – the East German experience", Discussion paper no. 97-33 (Zentrum für Europäische Wirtschaftsforshung).

Laffont, J. (1989), Fundamentals of public economics (MIT Press, Cambridge, MA).

LaLonde, R. (1984), "Evaluating the econometric evaluations of training programs with experimental data", Working paper no. 183 (Industrial Relations Section, Princeton University).

LaLonde, R. (1986), "Evaluating the econometric evaluations of training programs with experimental data", American Economic Review 76 (4): 604–620.

LaLonde, R. (1995), "The promise of public sector-sponsored training programs", Journal of Economic Perspectives 9 (2): 149–168.

LaLonde, R. and R. Maynard (1987), "How precise are evaluations of employment and training programs: evidence from a field experiment", Evaluation Review 11: 428–451.

Lancaster, T. (1990), Econometric analysis of transition data (Cambridge University Press for Econometric Society Monograph Series, Cambridge, UK).

Lechner, M. (1996), "An evaluation of public-sector-sponsored continuous vocational training programs in East Germany", Unpublished manuscript (Universität Mannheim).

Lechner, M. (1997), "Earnings and employment effects of continuous off-the-job training in East Germany after unification", Unpublished manuscript (Universität Mannheim).

Lee, L. (1983), "Generalized econometric models with selectivity", Econometrica 51 (2): 507–512.

Leigh, D. (1990), Does training work for displaced workers? (W.E. Upjohn Institute for Employment Research, Kalamazoo, MI).

Leigh, D. (1995), Assisting workers displaced by structural change (W.E. Upjohn Institute for Employment Research, Kalamazoo, MI).

Lewbel, A. (1998), "Semiparametric qualitative response model estimation with instrumental variables and unknown heteroskedasticity", Unpublished manuscript (Brandeis University).

Lewis, H.G. (1963) Unionism and relative wages (University of Chicago Press, Chicago, IL).

MaCurdy, T. (1982), "The use of time series processes to model the error structure of earnings in a longitudinal data analysis", Journal of Econometrics 18 (1): 83–114.

Main, B. (1985), "School leaver unemployment and the Youth Opportunities Programme in Scotland", Oxford Economic Papers 37 (3): 426–447.

Main, B. (1991), "The effects of the Youth Training Scheme on employment probability", Applied Economics 23 (2): 367–372.

Main, B. and D. Raffe (1983), "Determinants of employment and unemployment among school leavers: evidence from the 1979 survey of scottish school leavers", Scottish Journal of Political Economy 30 (1): 1–17.

Main, B. and M. Shelly (1990), "The effectiveness of the Youth Training Scheme as a manpower policy", Economica 57 (228): 495–514.

Mallar, C. (1978), "Alternative econometric procedures for program evaluations: illustrations from an evaluation of Job Corps", Proceedings of the American Statistical Association: 317–321.

Mallar, C., S. Kerachsky, C. Thornton and D. Long (1982), Evaluation of the economic impact of the Job Corps program: third follow-up report (Mathematica Policy, Princeton, NJ).

Manski C. (1995), The identification problem in the social sciences (Harvard University Press, Cambridge, MA)

Manski C. and S. Lerman (1977), "The estimation of choice probabilites from choice-based samples", Econometrica 45 (8): 1977–1988.

Manski, C. and D. McFadden (1981), "Alternative estimators and sample designs for discrete choice analysis", in: C. Manski and D. McFadden, eds., Structural analysis of discrete data with econometric applications (MIT Press, Cambridge, MA) pp. 1–50.

Masters, S. and R. Maynard (1981), The impact of supported work on long-term recipients of AFDC benefits (Manpower Demonstration Research Corporation, New York).

Matzkin, R. (1992), "Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models", Econometrica 60 (2): 239–270.

Matzkin, R. (1993), "Nonparametric identification and estimation of polychotomous choice models", Journal of Econometrics 58 (1,2): 137–168.

Maynard, R. (1980), The impact of supported work on young school dropouts (Manpower Demonstration Research Corporation, New York).

McLennan, A. (1991), "Binary stochastic choice", in: J. Chipman, D. McFadden and M. Richter, eds., Preferences, uncertainty and optimality: essays in honor of Leonid Hurwicz (Westview Press, Boulder, CO) pp. 187–202.

Mincer, J. (1962), "On-the-job training: costs, returns and some implications", Journal of Political Economy 70 (5): 50–79.

Mincer, J. (1993), "Investment in U.S. education and training", Discussion paper no. 671 (Columbia University).

Moffitt, R. (1992), "Evaluation methods for program entry effects", in: C. Manski and I. Garfinkel, eds., Evaluating welfare and training programs (Harvard University Press, Cambridge, MA).

National Commission for Employment Policy (1987), The Job Training Partnership Act (US Government Printing Office, Washington, DC).

Neyman, J. (1935), "Statistical problems in agricultural experiments", The Journal of the Royal Statistical Society 2 (2) (Suppl.): 107–180.

O'Connell, P. and F. McGinnity (1997), "What works, who works? The employment and earnings effects of active labour market programmes among young people in Ireland", Work, Employment and Society 11 (4): 639–661.

O'Higgins, N. (1994), "YTS, employment and sample selection bias", Oxford Economic Papers 46 (4): 605–628.

OECD (1993), "Active labour market policies: assessing macroeconomic and microeconomic effects", in: Employment outlook (OECD, Paris) pp. 39–67.

OECD (1996), Employment outlook (OECD, Paris).

Orr, L., H. Bloom, S. Bell, W. Lin, G. Cave and F. Doolittle (1994), "The National JTPA Study: impacts, benefits and costs of title II-A", Produced for the US Department of Labor under contract no. 99-6-0803-77-068 (Abt Associates, Bethesda, MD).

Park, N., W.C. Riddell and R. Power (1993) An evaluation of UI-sponsored training (Evaluation Branch, Human Resources Development Canada).

Payne, J., S. Lissenburg and M. White (1996), Employment training and employment action: an evaluation by the matched comparison method (Policy Studies Institute, London).

Perry, C., R. Anderson, R. Rowan and H. Northrupt (1975), The impact of government manpower programs (University of Pennsylvania Press, Philadelphia, PA).

Powell, J. (1994), "Estimation of semiparametric models", in: R. Engle and D. McFadden, eds., Handbook of econometrics, Vol. 4 (North-Holland, Amsterdam) pp. 2443–2521.

Puma, M. and N. Burstein (1994), "The national evaluation of the Food Stamp employment and training program", Journal of Policy Analysis and Management 13 (2): 311–330.

Quandt, R. (1972), "Methods for estimating switching regressions", Journal of the American Statistical Association 67 (338): 306–310.

Quandt, R. (1988), The economics of disequilibrium (Basil Blackwell, Oxford, UK).

Quint, J., B. Fink and S. Rowser (1994), New chance: interim findings on a comprehensive program for disadvantaged mothers and their children (Manpower Demonstration Research Corporation, New York).

Raaum, O. and H. Torp (1997), "Labour market training in Norway – effect on earnings", Report no. 46/97 (Stiftelsen for samfunns– ognaeringslivsforskning).

Rangarajan, A., J. Burghardt and A. Gordon (1992), Evaluation of the minority single parent demonstration, Vol. I: summary (Mathematica Policy Research, Princeton, NJ).

Rao, C.R. (1965), "On discrete distributions arising out of methods of ascertainment", in: G.P. Patil, ed., Classical and contagious discrete distributions (Statistical Publication Society, Calcutta).

Rao, C.R. (1986), "Weighted distributions", in: S. Feinberg, ed., A celebration of statistics (Springer-Verlag, Berlin).

Regner, H. (1996), "A nonexperimental evaluation of manpower training in Sweden", Unpublished manuscript (Stockholm University).

Regner, H. (1997), Training at the job and training for a new job: two Swedish studies (Swedish Institute for Social Research, Stockholm, Sweden).

Riccio, J., D. Friedlander and S. Freedman (1994), GAIN: benefits, costs and three-year impacts of a welfare-to-work program (Manpower Demonstration Research Corporation, New York).

Ridder, G. (1986), "An event history approach to the evaluation of training, recruitment and employment programmes", Journal of Applied Econometrics 11: 109–126.

Robins, J. (1989), "The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies", in: L. Sechrest, H. Freeman and A. Mulley, eds., Health service research methodology: a focus on AIDS (US Public Health Service, Washington, DC) pp. 113–159.

Robinson, P. (1996), "The role and limits of active labour market policy", Working paper RSC no. 96/27 (European Union Institute).

Rosenbaum, P. (1995) Observational studies (Springer-Verlag, Leipzig, Germany).

Rosenbaum, P. and D. Rubin (1983), "The central role of the propensity score in observational studies for causal effects", Biometrika 70 (1): 41–55.

Roy, A. (1951), "Some thoughts on the distribution of earnings", Oxford Economic Papers 3: 135–146.

Royden, H. (1968), Real analysis, 2nd edition (MacMillan Press, New York).

Rubin, D. (1974), "Estimating causal effects of treatments in randomized and non–randomized studies", Journal of Educational Psychology 66: 688–701.

Rubin, D. (1978), "Bayesian inference for causal effects: the role of randomization", Annals of Statistics 6 (1): 34–58.

Rubin, D. (1979), "Using multivariate matched sampling and regression adjustment to control bias in observational studies", Journal of the American Statistical Association 74: 318–328.

Sandell, S. and K. Rupp (1988), "Who is served in JTPA programs: patterns of participation and intergroup equity", US National Commission for Employment Policy RR-88-03.

Smith, J. (1992), "The JTPA selection process: a descriptive analysis", Unpublished manuscript (University of Chicago).

Smith, J. (1994), "A note on estimating the relative costs of experimental and non-experimental evaluations using cost data from the National JTPA Study", Unpublished manuscript (University of Chicago).

Smith, J. (1997a), "Measuring earnings dynamics among the poor: evidence from two samples of JTPA eligibles", Unpublished manuscript (University of Western Ontario).

Smith, J. (1997b), "Measuring earnings levels among the poor: evidence from two samples of JTPA eligibles", Unpublished manuscript (University of Western Ontario).

Smith, J. and F. Welch (1986), Closing the gap: forty years of economic progress for blacks (RAND, Santa Monica, CA).

Stromsdorfer, E., R. Boruch, H. Bloom, J. Gueron and F. Stafford (1985), Recommendations of the Job Training Longitudinal Survey Research Advisory Panel to the Office of Strategic Planning and Policy Development (US Department of Labor, Washington, DC).

Sudman, S. and N. Bradburn (1982), Asking questions (Jossey-Bass, San Francisco, CA).

Thierry, P. and M. Sollogoub (1995), "Les politiques francaises d'emploi en faveur des jeunes. Une evaluation econometrique", Revue-Economique 46 (3): 549–559.

Topel, R. and M. Ward (1992), "Job mobility and the careers of young men", Quarterly Journal of Economics 107: 439–480.

Torp, H., O. Raaum, E. Heraes and H. Goldstein (1993), "The first Norwegian experiment", in: K. Jensen and P.K. Madsen, eds., Measuring labour market measures (Ministry of Labour, Copenhagen, Denmark) pp. 97–140.

Trochim, W. (1984), Research design for program evaluation: the regression-discontinuity approach (Sage, Newbury Park, CA).

Trott, C. and J. Baj (1993), "An analysis of repeating in JTPA in Illinois", Report prepared for the Illinois Department of Commerce and Community Affairs (Northern Illinois University).

USGAO (1991), "Job Training Partnership Act: racial and gender disparities in services", Report no. GAO/HRD-91-148 (US General Accounting Office).

USGAO (1996), "Job Training Partnership Act: long-term earnings and employment outcomes", Report no. GAO/HEHE-96-40 (US General Accounting Office).

van der Klaauw, W. (1997), "A regression-discontinuity evaluation of the effect of financial aid offers on college enrollment", Unpublished manuscript (New York University).

Vytlacil, E. (1999), "Independence, monotonicity and latent variable models: an equivalence result", Unpublished manuscript (Univeristy of Chicago).

Weitzman, M. (1979), "Optimal search for the best alternative", Econometrica 47 (3): 641–654.

Westat (1981), "Continuous longitudinal manpower survey net impact report no. 1: Impact on 1997 earnings of new FY 1976 CETA enrollees in selected program activities", Report prepared for the US Department of Labor under contract no. 23-24-75-07.

Westat (1984), "Summary of net impact results", Report prepared for US Department of Labor under contract no. 23-24-75-07.

Westergard-Nielsen, N. (1993), "The effects of training: a fixed effect model", in: K. Jensen and P.K. Madsen, eds., Measuring labour market measures (Ministry of Labour, Copenhagen, Denmark) pp. 167–200.

White, M. and J. Lakey (1992), The Restart effect: does active labour market policy reduce unemployment? (Policy Studies Institute, London).

Whitfield, K. and C. Bourlakis (1991), "An empirical analysis of YTS, employment and earnings", Journal of Economic Studies 18 (1): 42–56.

Zweimuller, J. and R. Winter-Ebmer (1996), "Manpower training programmes and employment stability", Economica 63 (249): 113–130.