

SPRINGER BRIEFS IN ENERGY  
ENERGY ANALYSIS

Blair Fix

# Rethinking Economic Growth Theory From a Biophysical Perspective

Blair Fix

# Rethinking Economic Growth Theory From a Biophysical Perspective

 Springer

2015

# Acronyms

## Abbreviations

A	Technological Progress (Solow-Residual)
AG	Agriculture
$\alpha$	Capital's income share
$\beta$	Labor's income share
D	Total Debt
FF	Fossil Fuel
E	Energy
EROI	Energy Return on Investment
GDP	Gross Domestic Product
IN	Industry
J	Joule
K	Capital
KWH	Kilowatt-Hour
L	Labor
M	Mining Sector or Money Supply
n	Nominal
S	Services
SPP	Specific Purchasing Power
P	Unit-Price or Population <i>pc</i> Per Capita
V	Velocity of Money
Q	Quantity
R	Natural Resources
r	Real
$\rho$	Productivity
U	Useful Work
Y	Quantity of Final Production (unspecified unit)
Yn	Nominal GDP
Yr	Real GDP

## Data Sources

BEA	Bureau of Economic Analysis
BLS	Bureau of Labor Statistics
EIA	Energy Information Administration
HSUS	Historical Statistics of the United States

# Contents

<b>1</b>	<b>Introduction</b>	1
1.1	A Biophysical Perspective	1
1.2	Neoclassical Growth Theory	3
1.3	Methodological Approach: Relation to Existing Theory	5
1.3.1	The Biophysical Critique	6
1.3.2	The Functional Form Critique	7
1.3.3	The Measurement Critique	9
1.3.4	An Unorthodox Approach	11
1.4	Goals and Book Layout	13
	References	14
<b>2</b>	<b>Decoupling</b>	17
2.1	Natural Resources in Neoclassical Growth Theory	18
2.2	Measuring Decoupling	19
2.3	Problems with Real GDP	22
2.4	Decoupling: Fact or Artifact?	29
2.5	Decoupling in Theory and Reality: What Goes Wrong	38
2.6	Conclusions: Monetary Value as a Feedback	40
	References	42
<b>3</b>	<b>Distribution</b>	45
3.1	Measuring Growth	46
3.2	Functional Distribution	48
3.3	Debtor/Creditor Distribution	51
3.4	Personal Distribution	56
3.5	Conclusions: Distribution Matters	59
	References	60
<b>4</b>	<b>Institutional Size</b>	63
4.1	Institutions in Neoclassical Growth Theory	63
4.2	Theories of Large Institutions	67
4.2.1	Theories of the Firm	67

4.2.2	Theories of Government Size .....	69
4.3	Institutional Size and the Growth of Energy Consumption: Empirical Evidence .....	72
4.4	The Importance of Hierarchy .....	75
4.5	Conclusions: Size Matters .....	81
	References .....	82
<b>5</b>	<b>Labor Structure .....</b>	<b>85</b>
5.1	Labor in Neoclassical Growth Theory .....	86
5.2	Theories of Labor Structure Change .....	87
5.3	Three Hypotheses .....	90
5.4	Modelling Changes in Labor Structure .....	94
5.5	Labor Structure and Energy: Empirical Evidence .....	101
5.6	Conclusions: Labor Structure Matters .....	105
	References .....	106
<b>6</b>	<b>Toward a Biophysical Growth Theory .....</b>	<b>109</b>
6.1	The Problem with Neoclassical Growth Theory .....	109
6.2	A Biophysical Approach .....	111
6.3	Stylized Biophysical Facts .....	113
6.4	Conclusions .....	114
	References .....	116
	<b>Appendix US Sectors: Methodology and Sources .....</b>	<b>119</b>
	<b>Index .....</b>	<b>123</b>

# Chapter 1

## Introduction

*All theory depends on assumptions which are not quite true.  
That is what makes it theory.*

– Robert Solow (1956, p. 65)

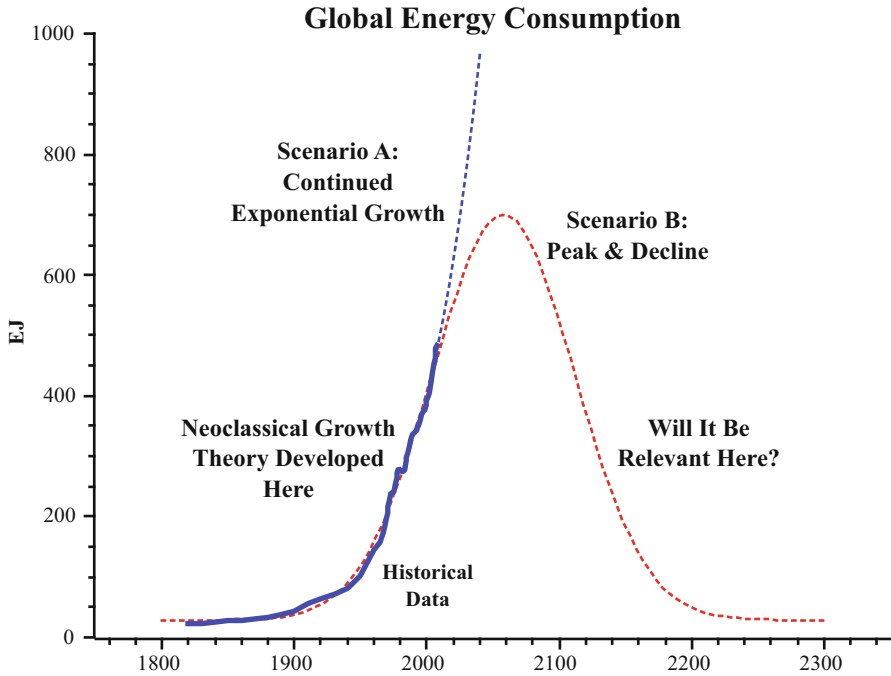
### 1.1 A Biophysical Perspective

The purpose of this book is to begin thinking about what a biophysical growth theory might look like. All theories must start by excluding from their scope the complexities of the real world that are deemed irrelevant. Such exclusion (or reductionism) makes scientific inquiry possible. For instance, prior to Isaac Newton, it was assumed that objects in motion tended to eventually come to a rest. However, by excluding the forces of friction from analysis, Newton was able to show that the opposite was true: objects in motion tended to stay in motion unless acted upon by an external force.

Yet excluding real-world complexities is a tricky business. If a theory excludes from its scope, the very forces that are causing the behavior in question, we have a problem. In this case, over-judicious simplification becomes a crutch: it makes mathematical formalization possible, but yields a model that has little to do with reality. Only through rigorous empirical investigation (and not through logical deduction), can one distinguish between good and bad assumptions.

Thus, the most appropriate starting point for a biophysical growth theory is to investigate the validity of assumptions made by existing growth theories. Since neo-classical growth theory is the dominant approach at the present time, this book empirically investigates *neoclassical assumptions*. The goal of this endeavor is to decide upon the essential elements of the growth process and to compile a set of stylized facts that a biophysical growth theory must successfully explain.

Not surprisingly, most existing growth theories are interested in explaining the phenomenon of sustained exponential *growth*. However, while exponential growth seems “normal” to the modern observer, it is a historical anomaly. The vast majority of human history has been characterized by stable equilibrium, not exponential growth. Figure 1.1 shows two scenarios for the future. Economists typically assume a future like *Scenario A*, with an indefinite continuation of exponential growth. Geologist M. King Hubbert, on the other hand, envisions a future like *Scenario B*, in which exponential growth is a “transient phenomenon in human history” (1993).



**Fig. 1.1** Two different visions of the future. (Sources: Modeled after Hall and Klitgaard (2012). Energy data for 1820–1960 from Smil (2010); 1965–2011 from BP Statistical Review of Energy, 2012. (Energy data includes oil, gas and coal))

Hubbert realized that the most important energy source used by industrial societies—fossil fuels—was finite, irreplaceable, and subject to inevitable depletion. He famously predicted the 1970 peak in US oil production, and made predictions that global oil production would peak around the year 2000. While recent evidence suggests that his prediction was slightly early (Hallock et al. 2014), the exact timing of these events is less important than the realization that they *will* occur, and that humanity’s energy future is far more likely to resemble *Scenario B* than *Scenario A*.

Hubbert also noted that, when we are far from the limits to growth, the unbounded exponential, Gaussian (bell-shape), and logistic (S-shape) curves are indistinguishable from one another. Thus, for a large portion of the pre-peak era, theories based on unbounded growth will accurately describe reality. It is only as we near the peak that theories of unbounded growth will fail. Even though *Scenario B* is only hypothetical, note that it departs from *Scenario A* in the early twenty-first century. Interestingly, at the present time the sluggish growth of many countries is perplexing many mainstream economists. Biophysical scholars such as Hall and Klitgaard (2012) suggest that we are nearing the limits to growth. Given this reality, a growth theory that will remain relevant in the future must explain both growth and forced degrowth.

When testing neoclassical assumptions, it is important to recognize that neoclassical economists are often acutely aware that their assumptions are “not quite true.” Indeed, in his 1956 paper outlining the foundations of neoclassical growth theory, Robert Solow acknowledges this in his first sentence (see opening quotation).<sup>1</sup> However, Solow and other neoclassical theorists typically regard such untrue assumptions as “innocuous”—a necessary part of constructing a theory. This book is concerned with the possibility that assumptions that are perceived as *innocuously* untrue (by neoclassical theorists) are actually *insidiously* untrue. This will be the case if neoclassical assumptions exclude from analysis the very phenomena that are most central to the growth process. The empirical work conducted in this book aims to find out whether or not this is the case.

## 1.2 Neoclassical Growth Theory

There are many variants of neoclassical growth theory, but this book focuses on the Solow–Swan model, the “workhorse model of macroeconomics” (Acemoglu 2008, p. 26). This model bears the name of its two main contributors, Robert Solow (1956, 1957) and Trevor Swan (1956). In his textbook, *Introduction to Modern Economic Growth*, Daron Acemoglu writes that this model has “shaped the way [neoclassical economists] approach not only economic growth but also the entire field of macroeconomics” (ibid, p. 26).

While the Solow–Swan model is now the canonical approach used in macroeconomics, it is important to recognize that it came into existence as a critique of an earlier post-Keynesian model of growth put forward by Harrod (1939) and Domar (1946). As Hagemann notes, the Harrod–Domar model was characterized by a “secular instability problem, which [was] thoroughly Keynesian in spirit and a kind of extension of the unemployment problem to the long run” (2009, p. 72). The Harrod–Domar model emphasized the basic instability of capitalist growth—a topic that was freshly in mind when the model was developed (shortly after the Great Depression).

Solow began his model by accepting “all the Harrod–Domar assumptions except that of fixed proportions” (1956, p. 66). By doing away with fixed proportions and allowing substitution between the factors of production (labor and capital), Solow was able to create a model in which *stability*, rather than instability, was the norm. This stability was a highly desirable feature for a neoclassical model, because it meshed well with the neoclassical tenet that the free market is inherently self-regulating. Hall

---

<sup>1</sup> Interestingly, Solow was referring to assumptions made by the Harrod–Domar model (not his own assumptions). In the same paragraph, he continues: “When the results of a theory seem to flow specifically from a special crucial assumption, then if the assumption is dubious, the results are suspect” (1956, p. 65). While his aim is the Harrod–Domar assumption of fixed factor proportions, one can also argue (as I do in this book) that the results of the Solow–Swan model are also dependent on “dubious” assumptions.



and Klitgaard note that this allowed Solow to “turn a social problem [instability] into a technical one, and [maintain] the neoclassical ideal of self-regulating markets over the long term” (2012, p. 127). Thus, we can see that the assumptions underlying neoclassical growth theory are not random; rather, they are chosen specifically to produce a model that is consistent with neoclassical principles.

The Solow–Swan model assumes an economy populated by homogeneous households and firms, such that each can be represented in aggregate by a single household and a single firm. Households own the two factors of production (labor and capital) and rent them to firms for use in production. Investment (and thus, capital accumulation) is a function of the household savings rate, which is assumed to be constant. The single sector economy produces one unique good that is both consumed and used as capital for further production.

At the core of this model is a neoclassical aggregate production function, which usually takes Cobb–Douglas form:<sup>2</sup>

$$Y = AL^{\beta}K^{\alpha} \quad (1.1)$$

Here  $Y$  refers to the annual quantity of final output,  $L$  the quantity of labor input,  $K$  the quantity of capital input, and  $A$  a dimensionless multiplier of the production function (sometimes called “total factor productivity”) considered to represent technological change. The Cobb–Douglas production function is a simple method of mapping the “factors of production” (i.e., inputs) onto output in a way that satisfies the requirements of neoclassical theory. Curiously, it does not include any consideration of the necessary requirements of energy or materials.

A central tenet of neoclassical distribution theory is that add, each factor of production should receive income in proportion to its marginal productivity. As a result of this theory, the exponents in the production function are predicted to equal the income share of each factor. Thus,  $\alpha$  represents capital’s share of national income and  $\beta$  labor’s share. Both sum to one, which guarantees “constant returns to scale,” meaning changes in input are linearly related to changes in output. This is important because a central “proof” of the marginal productivity theory of distribution (the application of Euler’s theorem) only works under the assumption of constant returns to scale (Robinson 1934). Furthermore, the income shares of labor and capital are assumed to be constant over time.

In order to maintain simplicity, the Solow–Swan model makes numerous assumptions (Table 1.1) that are clearly untrue. Yet the hope is that such assumptions merely remove complexities of the real world that are not relevant to the growth process. The result is a highly idealized model, that, while untrue in the strict sense, potentially gives true insight into the core aspects of growth. The kernel of the Solow–Swan model, which has remained the central tenet of all subsequent neoclassical growth theories, is that growth is a stable process that is the result of capital accumulation and

---

<sup>2</sup> This functional form owes its name to the work of Cobb and Douglas (1928), but the original use of the equation dates back to the work of Wicksell (1911)[1901].

**Table 1.1** Explicit assumptions made by the Solow–Swan model. (Sources: Acemoglu (2008); Solow 1956)

1.	Constant depreciation of capital	9.	Homogeneous households/firms
2.	Closed economy (no trade)	10.	No debt financing
3.	Constant returns to scale	11.	No government
4.	Constant savings rate	12.	Perfect competition
5.	Diminishing marginal productivity of K, L	13.	Profit maximization
6.	Incomes proportional to marginal productivity	14.	Single commodity produced
7.	Full employment	15.	Single economic sector
8.	Free technology		

**Table 1.2** Implicit assumptions made by the Solow–Swan model

1.	Economic output can become decoupled from energy inputs
2.	Economic distribution is unrelated to growth
3.	Large institutions are not important for growth
4.	Labor force structure is not important for growth

technological progress (increases in the labor supply play a relatively insignificant role).

The validity of this kernel, however, depends on the *marginal* role played by the complexities that are assumed away. However, if these complexities are not marginal, but rather, *central* to the growth process, results derived from the neoclassical growth model may be grossly misleading. Therefore, it is of the utmost importance to rigorously test such assumptions. Unfortunately, neoclassical theory, for the most part, fails to do so.

This book does not test the explicit assumptions displayed in Table 1.1, but rather, the implicit assumptions displayed in Table 1.2. All four of these implicit assumptions follow logically from the explicit assumptions made by the Solow–Swan model. By investigating these implicit assumptions, we are effectively testing whether or not growth is as simple as neoclassical theory implies. If these four assumptions turn out to be justified, then neoclassical growth theory stands on solid ground. However, if these assumptions are not justified, then we must conclude that neoclassical growth theory ignores processes that are essential to growth.

### 1.3 Methodological Approach: Relation to Existing Theory

The research that I have conducted in this book is based on a deep-seated dissatisfaction with the state of conventional growth theory. My dissatisfaction is not unique. Indeed, there is a long lineage of scholarship that has critiqued neoclassical growth theory from a variety of perspectives. Below, I review a few of these schools of thought and situate my own approach within this scholastic backdrop.

### 1.3.1 *The Biophysical Critique*

We begin with critiques of neoclassical growth theory that come from a biophysical perspective. Nicholas Georgescu-Roegen (1971) argued that the economy is a biophysical entity that is subject to the laws of thermodynamics. However, Hall et al. note that neoclassical theory “pays only marginal attention to the first and second laws of thermodynamics” (2001, p. 664). Herman Daly argues that the absence of resources in the typical neoclassical production function is like a recipe that calls for “making a cake with only the cook and his kitchen” (1997, p. 261).

Some scholars have pursued ways of reforming neoclassical growth theory by integrating natural resources into the Cobb–Douglas production function (Stiglitz 1974). However, one of the difficulties with this approach is that certain neoclassical conditions conspire to make resources unimportant. Neoclassical distribution theory predicts that the exponents in the production function (the “elasticities of output”) should be equal to each factor of production’s share of income. But, since resource owners typically get paid only a tiny fraction of the national product, resource exponents must be small. Thus, if one adopts strict neoclassical conditions, resources will inherently be given little significance (Ayres and Warr 2005).

Because of this problem, scholars who add natural resources to production functions have often had to abandon the neoclassical connection between output elasticities and income shares. Ayres and Warr note that “including a fossil energy flow proxy in the neoclassical production function, without any constraint on factor share, seems to account for economic growth quite accurately, at least for limited time periods, without any exogenous time-dependent term” (2005, p. 182). Examples of this “reformed” production function approach include Hannon and Joyce (1981), Kümmel (1982, 1989), Cleveland et al. (1984), Kümmel et al. (1985, 2000), Kaufman (1992), and Beaudreau (1998).

Along with the addition of energy as a factor of production (and the abandonment of the factor share assumption) it has become common to replace the Cobb–Douglas production function with a linear-exponential (“LINEX”) production function (Kümmel 1989; Hall et al. 2001; Ayres and Warr 2005, 2009). In the work of Ayres and Warr, the LINEX production function takes the following form:

$$Y = AU \exp \left( \alpha \frac{L}{U} - \beta \frac{U + L}{K} \right) \quad (1.2)$$

Here  $\alpha$  is a constant and  $U$  represents “useful work”, while the remainder of the variables are the same as in the Cobb–Douglas function (Eq. 1.1). Note, however, that  $\alpha$  and  $\beta$  no longer represent capital and labor’s share of income.

This approach has met with great empirical success. In my opinion, the most significant achievement of the biophysical perspective has been to emphasize the importance of energy to the growth process. My own work accepts this importance as a central tenet. As such, when I say that I adopt a “biophysical perspective”, this means that I focus on energy flows, and that I heed Georgescu-Roegen’s call to make economic theory compatible with the laws of thermodynamics.

However, I believe that the “reformed” production function approach does not go far enough in its critique of neoclassical growth theory. As Shaikh notes, “The aggregate production function is a fundamental[ly] neoclassical construct” (2005, p. 447). Indeed, Nobel Prize winning economist Edward Prescott states that “the neoclassical production function is the cornerstone of the theory and is used in virtually all applied aggregate analyses” (1998, p. 532).

As discussed below, there are major difficulties with the use of production functions. My worry is that by adopting a production function approach, one implicitly accepts some of the neoclassical tenets that one is testing.

### 1.3.2 The Functional Form Critique

The functional form critique highlights fundamental problems with the basic form of the Cobb–Douglas production function. For instance, Mayumi, Giampietro, and Ramos-Martin critique Cobb–Douglas functions as being part of a “curve-fitting fetishism in economics” (2012, p. 26). They argue that neoclassical production functions violate dimensional rules:

Suppose that  $K$ ,  $L$ , and  $Y$  are represented in terms of the US dollar. Since  $\alpha + (1 - \alpha) = 1$ , the dimension of the left-hand side, the US dollar, is compatible with that of the right-hand side as a whole if  $A$  is a dimensionless pure number. However, each term on the right-hand side (i.e.,  $K$  and  $L$ ) does not make any sense unless  $\alpha = 0$  or  $1$ . Suppose  $\alpha = 1/2$ ; is there any operational meaning of  $\sqrt{100 \text{ \$US}}$ , for example? Thus we are at a loss to understand the true reason why the Cobb–Douglas specification is often used in economic science. (ibid, p. 26–27)

Perhaps the main reason for the continued use of production functions is that they appear to “work.” By using such a function, one can make falsifiable predictions that are confirmed by empirical evidence. However, these predictions are usually not about growth itself, but rather, about the *distribution* of income. Neoclassical growth theory predicts that the exponents  $\alpha$  and  $\beta$  should be equivalent to each factor’s share of income. Neoclassical theory defines the ‘output elasticity’ with respect to labor (the change of output per change of labor input) as the marginal product of labor ( $\frac{\partial Y}{\partial L}$ ), times the amount of labor per unit of output ( $L/Y$ ):

$$\text{Output Elasticity} = \frac{\partial Y}{\partial L} \cdot \frac{L}{Y} = \beta \quad (1.3)$$

Carrying out this operation on Eq. 1.1 yields the result that the exponent  $\beta$  is the output elasticity with respect to labor. Neoclassical distribution theory predicts that the wage rate ( $w$ ) should be equal to the marginal product of labor:

$$w = \frac{\partial Y}{\partial L} \quad (1.4)$$

Substituting Eq. 1.4 into 1.3 yields:

$$\beta = \frac{wL}{Y} \quad (1.5)$$

Thus, neoclassical theory predicts that  $\beta$  is equal to labor's share of output ( $wL/Y$ ). Carrying out a similar operation shows that  $\alpha$  is predicted to be capital's share of income, and that  $\alpha$  and  $\beta$  should sum to one. Using regressions of historical data, economists have empirically estimated these exponents. Walters notes the results:

First, the sum of the coefficients usually approximates closely to unity. The linearity of the production function seems to be a remarkably consistent finding between one country and another. The second important result is the agreement between the labour exponent and the share of wages in the value of output. These two findings have been interpreted as confirmation that the aggregate production function has constant returns to scale and that the marginal productivity of labour is equal to the wage rate. (1963, p. 27)

For many economists, these strong empirical results justify the use of aggregate production functions. However, Franklin Fisher argues that the empirical robustness of the Cobb–Douglas production function is a paradox. Fisher has focused specifically on the conditions under which it is possible to move from firm-specific production functions to an economy-wide *aggregate* production function. His findings are not reassuring. He notes that “the conditions under which . . . a technologically diverse economy can be represented by an aggregate production function are *far too stringent to be believable*” (1971, p. 305, emphasis added). In order to understand why aggregate production functions appear to “work” when they should not, Fisher created a model in which the conditions for aggregation were purposefully *not* met. Yet, to his surprise, as long as labor and capital shares were constant, he found that a Cobb–Douglas function could be used to accurately fit aggregate data.

McCombie (2001) argues that the origin of this paradox stems from the use of monetary data for empirical work. Neoclassical growth theory is derived from a one-commodity thought experiment, in which both inputs and outputs can be measured in *physical* terms. However, the real-world application of the theory requires that measurement be based on *monetary value* (since inputs and outputs are heterogeneous and cannot be aggregated in physical units). This resort to monetary value means that the data used for analysis is generated using the national accounts *identity*:

$$Y \equiv wL + rK \quad (1.6)$$

The national accounts are an accounting system used by government agencies to measure, among other things, the “national wealth”  $Y$  (i.e., total real value added). Equation 1.6 is one way of *defining* the national wealth. That is, total value-added ( $Y$ ) is *defined* as the sum of all individual “real” incomes. Grouping income by functional type, this becomes the income of all workers (the average real wage rate  $w$  times the number of workers  $L$ ) plus the income of all capitalists (the rate of profit  $r$  times the real value of the capital stock  $K$ ). Remember that this equation is definitionally true for any conceivable value of  $w$ ,  $L$ ,  $r$ , and  $K$ .

Anwar Shaikh (1974) resolved Fisher's paradox by showing that through algebraic manipulation, the national accounts identity could be rewritten into a form that is *identical* to a Cobb–Douglas production function (Eq. 1.7). All that was needed was the assumption that both factor shares and the growth rates of wages and profits remained constant.

$$Y = AL^\beta K^{1-\beta} \quad (1.7)$$

Here,  $\beta$  is labor's share (Eq. 1.8), while  $A$  is a smooth exponential growth function (Eq. 1.9) whose parameter  $\psi$  is the weighted average of the growth rates of the wage and profit rates (Eq. 1.10). For an explanation of the steps involved in this algebraic transformation, see Shaikh (1974) and/or Felipe and Holtz (2001).

$$\beta = wL/Y \quad (1.8)$$

$$A = A_0 e^{\psi t} \quad (1.9)$$

$$\psi = \beta \hat{w} + (1 - \beta) \hat{r} \quad (1.10)$$

Note that Eq. 1.7 through 1.10 contain nothing that does not follow directly from the national accounts identity: they are *definitionally* true. To emphasize this point, Shaikh showed that even a production relation that traced the word “HUMBUG” (with capita per head on the horizontal axis and output per head on the vertical axis) could be successfully fitted with a Cobb–Douglas production function using the method provided by Solow 1957.

The key assumption used to derive Eq. 1.7 through 1.10 was that both factor shares and the growth rates of wages and profits remain constant. If this holds approximately true in the real world (and it often does), production function estimates will uncover the underlying identity. Based on Shaikh's results, McCombie concludes:

... [T]he underlying accounting identity ... prevents the empirical testing of the production function. Thus, it invalidates the traditional defense of the use of the production function, which is that, in spite of all the reservations about its underlying theoretical rationale, ... it is still legitimate to use production functions because of their good predictive ability. (2001, p. 613)

Edwin Burmeister, noting his discontent with aggregate production functions, makes the following suggestion:

Perhaps for the purpose of answering many macroeconomic questions ... we should disregard the concept of a production function at the macroeconomic level. The economist who succeeds in finding a suitable replacement will be a prime candidate for a future Nobel prize (Burmeister 2008, pp. 427–428, cited in Felipe and Fisher 2003, p. 247)

I take the functional form critique seriously. As such, I do not use production functions in this book.

### 1.3.3 The Measurement Critique

The “measurement critique” highlights epistemological difficulties surrounding the measurement of production function inputs and outputs. The starting point for neo-classical growth theory is to assume an economy in which both capital input and final output consist of a *single, unchanging commodity*. In formulating his model, Solow writes: “There is only one commodity, output as a whole, whose rate of production is designated  $Y(t)$ . Thus we can speak unambiguously of the community's real income” (1956, p. 66).

In such a model, inputs and outputs are well defined. Both capital and output consist of the same single good. However, in the real world, there is a myriad of different commodities that change over time. The question is, when moving from a single-commodity model to the real world, can we still “speak unambiguously” of inputs and outputs? This question was at the heart of what became known as the “Cambridge Capital Controversy,” a debate between Cambridge (UK) economists Joan Robinson and Piero Sraffa, and Cambridge (US) economists such as Paul Samuelson and Robert Solow.

Joan Robinson (1953) set off the debate by asking a very simple question: in what unit is “capital” to be measured? Nitzan and Bichler (2009) note that in neoclassical theory, capital exists as a duality, having both a physical and financial magnitude. While only the financial magnitude can be measured, it is hoped (by neoclassical economists) that the financial magnitude will, over the long run, somehow reflect the underlying “physical” quantity of capital.

However, by measuring capital in terms of its monetary value, many problems arise. Cohen and Harcourt summarize one aspect of the problem:

Heterogeneous capital goods cannot be measured and aggregated in physical units; instead, capital valuation must be used, as Wicksell (1911) told us long ago. Their value can be measured either as the cost of production, which takes time, or the present value of the future output stream they produce. In either case, since the measure involves time, it presumes a rate of interest—which, in the simple model, is determined in a one-way manner by the quantity of capital. This additional circularity, or interdependence, causes Wicksell effects. Wicksell effects involve *changes in the value of the capital stock associated with different interest rates*, arising from either inventory revaluations of the same physical stock due to new capital goods prices (price Wicksell effects) or differences in the physical stock of capital goods (real Wicksell effects). (2003, pp. 201–202) [emphasis added]

Charles Hulten concludes that “the methods used for imputing the price and quantity of capital input are strongly neoclassical” (2009, p. 49). This raises important epistemological issues for those seeking to test neoclassical theory. A cardinal rule for the non-biased test of a theory is that one cannot use metrics that themselves are influenced by the theory that one is testing. If, as Hulten states, the measurement of capital is “strongly neoclassical”, then the use of such a metric in a test of neoclassical theory poses significant epistemological issues.

While the Cambridge controversy revolved around problems with the aggregation of capital, Felipe and Fisher note that “there exist equally important labor and output aggregation problems” (2003, p. 228). Here I focus on problems with aggregating output. Just like capital, output must be measured in monetary rather than physical terms. This measurement is conceptually very simple: nominal gross domestic product (GDP) is deflated by a price index that captures the change in the average price level.

However, as discussed extensively in Chap. 2, this process is not as straightforward as it sounds. One key problem involves changes in relative prices. Francis Edgeworth (1887) notes that:

If one great group of commodities varies pretty uniformly in one direction, and another in a different direction (or even in the same direction but in a markedly different degree), then the task of restoring the level of prices can no longer be regarded as a *purely objective* . . . (cited in Vining and Elwertowski 1976, p. 699)[emphasis added]

Landefeld and Parker (1997) explain aspects of the problem:

... because relative prices and associated patterns of purchases change over time, ... GDP growth will be quite sensitive to the choice of the base year, and a shift in the base year often has a significant impact on the measured growth rates.

... In periods such as the energy crisis of 1973–75, relative price and consumption patterns can change rapidly, and significant bias can creep into fixed-weighted measures even during periods close to the base period.

If, as Edgeworth argues, there are irreducibly subjective elements to the calculation of price indices, and thus, to the calculation of real GDP, we must take such claims seriously, for they have a direct impact on the quality of empirical research. This subjectivity means that if different methodological choices are adopted, the resulting output metric might be different. Giampietro et al. call this the “epistemological predicament associated with purposive quantitative analysis ... the observer always affects what is observed when defining the descriptive domain” (2006, p. 307). As shown in Sect. 2.3, different methodologies for calculating inflation can lead to *wildly* different estimations of real GDP.

A further problem with real GDP is that, like capital, it is not clear that its calculation is sufficiently independent from neoclassical theory to allow a nonbiased test of such theory. Jonathan Nitzan writes:

In a certain fundamental sense ... our data on how prices and output change may not be sufficiently independent from our views on why they change and [this] inherent subjectivity must be recognized. Within the present historical epoch, the predisposition of price and quantity data toward the neoclassical economic outlook means that these data may not be altogether suitable to test the neoclassical outlook against competing frameworks. (1992, p. 158)

I take the measurement critique very seriously, especially the possibility that the conventional measure of output (real GDP) is circularly dependent on neoclassical theory. As such, I adopt an empirical methodology that does not rely on real GDP.

### 1.3.4 *An Unorthodox Approach*

My objective in this book is to test the assumptions made by neoclassical growth theory in a way that is *completely* independent of neoclassical theory itself. In order to maintain this independence, I make the following radical methodological decisions: (1) I abandon the use of production functions; and (2) I abandon the use of real GDP.

This second decision is particularly unorthodox. After all, real GDP is *the* growth metric used by economists. My decision to abandon real GDP is based on the contributions of Nitzan and Bichler (2009) to capital theory. Nitzan and Bichler critique the “real/nominal” duality that underpins many aspects of economics. This duality assumes that monetary value (the “nominal” sphere) somehow represents an underlying physical magnitude (the “real” sphere).

Thus, in capital theory, the value of financial capital is assumed to represent the scale of the physical capital stock. Similarly, in growth theory, it is assumed



that the *value* of output (“nominal” GDP) can be transformed into the *quantity* of output (“real” GDP). Yet, Nitzan and Bichler argue that there are insurmountable epistemological problems with this duality. Their contribution to capital theory is to reject the real/nominal duality by insisting that capital is a purely financial magnitude. In this book, I take Nitzan and Bichler’s capital theory insights and apply them to growth theory. Thus, I retain the use of nominal GDP (a purely financial magnitude), but abandon the use of real GDP.

But one might ask—is it even possible to conduct empirical research about growth without using real GDP? I argue that, not only is it *possible*, it is highly *desirable* (from a biophysical perspective). I assert that a biophysical growth theory should be concerned with *biophysical* scale, not monetary scale.

Before discussing my chosen measure of biophysical scale, let us take a step back and consider the following thought experiment. Suppose that one day humans encounter an alien race with advanced technology. However, this technology (and other aspects of their economy) are very different than anything with which we are familiar. Given this situation, how would we measure the “scale” of their economy in order to compare it to our own? As we do not understand their economy, we cannot speak of “output” in any meaningful sense. However, since the laws of physics and chemistry presumably apply the same to these aliens as they do to us, we could quite easily measure the raw material *throughput* of their economy. If we insist on a biophysical measure of scale, then there are really only two ways to measure this material flow—in terms of mass or in terms of energy. Both will yield different results for the “scale” of the economy.

Biophysical scholars have recently begun to measure economies in biophysical terms. For instance Gierlinger and Krausmann (2012) measure the historical scale of the USA in terms of mass flows. As such, I could use mass flow as my measure of biophysical scale (however, Gierlinger and Krausmann have not, to my knowledge, made their data public). Giampetro et al. (2012) explore the properties of an indicator they call “bio-economic pressure”—the total energy consumption of a society divided by the human activity in the industrial (and agricultural) sector. In the spirit of this exploration, I make the decision to adopt *energy throughput* as a biophysical measure of scale. When relevant data is available, I use Ayres and Warr’s (2005) estimates of useful work.

In order to avoid confusion, I should be clear that this does *not* mean that I am using energy use as a “proxy” for output. Energy is an *input* to the economy. I adopt it as a measure of biophysical scale because it is a particularly *important* input, but I make no claim that energy has a monopoly on measurements of scale. Nor do I claim that energy is a “better” measure of scale than is real GDP—it is simply different. Furthermore, I do *not* claim that energy use and real GDP are equivalent. Such a claim would make my work redundant: if the two were the same, then there would be no reason for adopting energy as a measure of scale. The whole point of my approach is to use a metric of biophysical scale that departs *significantly* from the orthodox output metric of real GDP. By doing so, one opens the door for empirical results that would not otherwise be possible if only real GDP were used for analysis.

When testing neoclassical growth theory, I am forced to adopt a slight of hand. As the theory is explicit that “economic growth” means the growth of the “quantity of output”, my approach is to transform neoclassical assumptions about “output” into assumptions regarding “biophysical scale”. One may protest that such slight of hand is unjustified. However, the reality is that neoclassical growth theory *cannot* be tested on its own terms. The neoclassical growth model is explicitly derived from the assumption of a one-commodity economy in which output is defined in *physical terms*. However, the heterogeneity of actual output means that such physical measurement is impossible.

Thus, neoclassical economists also adopt a slight of hand when testing their theory. When applying their model to the real world, economists drop the pretence that output can be measured in physical terms. Instead, the physical output term  $Y$  is quietly replaced with an *index* of output measured in terms of *monetary value* (thus assuming the validity of the real/nominal duality). As I find this neoclassical slight of hand untenable, I choose to substitute my own. I believe that the need for a new growth theory is so great that this highly unorthodox methodology is justified.

The approach adopted in this book requires a leap of faith on the part of the reader who is trained in conventional economics. If one is not prepared to make such a leap, then read no further. However, if one is interested in what happens when implicit neoclassical assumptions are tested in a highly unorthodox manner, then continue reading. Due to the groundbreaking nature of this approach, all results should be considered preliminary. For the reader who is willing to make a leap of faith, I ask only that one judge my approach on the robustness of the empirical evidence.

## 1.4 Goals and Book Layout

The goal of this book is twofold: (1) to contribute to the creation of a new biophysical growth theory by pointing out some of the major problems of conventional (neoclassical) growth theory; and (2) to outline empirical components that I believe should form part of the basis for a biophysical growth theory.

In the following chapters, each of the four implicit neoclassical assumptions in Table 1.2 is tested empirically. Chapter 2 focuses on decoupling, Chap. 3 on distribution, Chap. 4 on institutional size, and Chap. 5 on labor structure. In all four empirical studies, the evidence demonstrates that neoclassical growth theory makes assumptions that exclude from analysis some of the most *fundamental* aspects of the growth process. In an attempt to crystallize the important elements of a future biophysical growth theory, Chap. 6 concludes by formulating a set of stylized facts that such a theory must successfully explain.

## References

- Acemoglu D (2008) Introduction to modern economic growth. Princeton University Press, Princeton
- Ayres R, Warr B (2005) Accounting for growth: the role of physical work. *Struct Chang Econ Dyn* 16(2):181–209
- Ayres R, Warr B (2009) The economic growth engine: how energy and work drive material prosperity. Edward Elgar Pub, Northampton
- Beaudreau BC (1998) Energy and organization: growth and distribution re-examined. Greenwood Publishing Group, Westwood
- British Petroleum (2012) BP statistical review of world energy. British Petroleum, London
- Burmeister E (2008) Comment. In: Usher D (ed) The measurement of capital. University of Chicago Press, Chicago, pp 420–431
- Cleveland C, Costanza R, Hall C, Kaufmann R (1984) Energy and the US economy: a biophysical perspective. *Science* 225(4665):890–897
- Cobb CW, Douglas PH (1928) A theory of production. *Am Econ Rev* 18(1):139–165
- Cohen AJ, Harcourt GC (2003) Retrospectives: whatever happened to the Cambridge capital theory controversies? *J Econ Perspect* 17(1):199–214
- Daly H (1997) Georgescu-Roegen versus Solow/Stiglitz. *Ecol Econ* 22(3):261–266
- Domar ED (1946) Capital expansion, rate of growth, and employment. *Econom, J Econom Soc* 14(2):137–147
- Edgeworth FY (1887) Measurement of change in value of money i. First Memorandum presented to the British Association for the Advancement of Science Reprinted in his Papers Relating to Political Economy 1:198–259
- Felipe J, Fisher FM (2003) Aggregation in production functions: what applied economists should know. *Metroeconomica* 54(2–3):208–262
- Felipe J, Holz CA (2001) Why do aggregate production functions work? Fisher’s simulations, Shaikh’s identity and some new results. *Int Rev Appl Econ* 15(3):261–285
- Fisher FM (1971) Aggregate production functions and the explanation of wages: a simulation experiment. *The Review of Economics and Statistics* 53(4):305–325
- Georgescu-Roegen N (1971) The entropy law and the economic process. Harvard University Press, Cambridge
- Giampietro M, Allen TFH, Mayumi K (2006) The epistemological predicament associated with purposive quantitative analysis. *Ecol Complex* 3(4):307–327
- Giampietro M, Mayumi K, Sorman A (2012) The metabolic pattern of societies: where economists fall short. Routledge, New York
- Gierlinger S, Krausmann F (2012) The physical economy of the United States of America. *J Ind Ecol* 16(3):365–377
- Hagemann H (2009) Solow’s 1956 contribution in the context of the Harrod-Domar model. *Hist Polit Econ* 41(Suppl 1):67–87
- Hall C, Klitgaard K (2012) Energy and the wealth of nations: understanding the biophysical economy. Springer, New York
- Hall C, Lindenberger D, Kümmel R, Kroeger T, Eichhorn W (2001) The need to reintegrate the natural sciences with economics. *BioScience* 51(8):663–673
- Hallock JL, Wu W, Hall CA, Jefferson M (2014) Forecasting the limits to the availability and diversity of global conventional oil supply: validation. *Energy* 64:130–153
- Hannon B, Joyce J (1981) Energy and technical progress. *Energy* 6(2):187–195
- Harrod RF (1939) An essay in dynamic theory. *Econ J* 49(193):14–33
- Hubbert MK (1993) Exponential growth as a transient phenomenon in human history. In: Daly H, Townsend K (eds) Valuing the earth: economics, ecology ethics. MIT Press, Cambridge, pp 113–126
- Hulten CR (2009) Growth accounting. Tech. rep., National Bureau of Economic Research. <http://www.nber.org/papers/w15341>. Accessed 23 May 2014

- Kaufmann RK (1992) A biophysical analysis of the energy/real GDP ratio: implications for substitution and technical change. *Ecol Econ* 6(1):35–56
- Kümmel R (1982) The impact of energy on industrial growth. *Energy* 7(2):189–203
- Kümmel R (1989) Energy as a factor of production and entropy as a pollution indicator in macroeconomic modelling. *Ecol Econ* 1(2):161–180
- Kümmel R, Strassl W, Gossner A, Eichhorn W (1985) Technical progress and energy dependent production functions. *J Econ* 45(3):285–311
- Kümmel R, Lindenberger D, Eichhorn W (2000) The productive power of energy and economic evolution. *Indian J Appl Econ* 8(2):1–26
- Landefeld JS, Parker RP (1997) BEA's chain indexes, time series, and measures of long-term economic growth. *Surv Curr Bus* 77(5):58–68
- Mayumi K, Giampietro M, Ramos-Martin J (2012) Reconsideration of dimensions and curve-fitting practice in view of Georgescu-Roegen's epistemology in economics. *J Econ Forecast* 15(4):17–35
- McCombie JSL (2001) What does the aggregate production function show? Further thoughts on Solow's "second thoughts on growth theory". *J Post Keynes Econ* 23(4):589–615
- Nitzan J (1992) Inflation as restructuring. a theoretical and empirical account of the US experience. PhD thesis, McGill University
- Nitzan J, Bichler S (2009) Capital as power: a study of order and creorder. Routledge, New York
- Prescott EC (1998) Lawrence R. Klein lecture 1997. Needed: a theory of total factor productivity. *Int Econ Rev* 39(3):525–551
- Robinson J (1934) Euler's theorem and the problem of distribution. *Econ J* 44(175):398–414
- Robinson J (1953) The production function and the theory of capital. *Rev Econ Stud* 21(2):81–106
- Shaikh A (1974) Laws of production and laws of algebra: the humbug production function. *Rev Econ Stat* 56(1):115–120
- Shaikh A (2005) Nonlinear dynamics and pseudo-production functions. *East Econ J* 31(3):447–466
- Smil V (2010) Energy transitions: history, requirements, prospects. Praeger Publishers, Santa Barbara
- Solow RM (1956) A contribution to the theory of economic growth. *Q J Econ* 70(1):65–94
- Solow RM (1957) Technical change and the aggregate production function. *Rev Econ Stat* 39(3):312–320
- Stiglitz J (1974) Growth with exhaustible natural resources: efficient and optimal growth paths. *Rev Econ Stud* 41(Symposium on the Economics of Exhaustible Resources):123–137
- Swan TW (1956) Economic growth and capital accumulation. *Econ Rec* 32(2):334–361
- Vining DR, Elwertowski TC (1976) The relationship between relative prices and the general price level. *Am Econ Rev* 66(4):699–708
- Walters AA (1963) Production and cost functions: an econometric survey. *Econom: J Econom Soc* 31(1/2):1–66
- Wicksell K (1911) Lectures on political economy, Vol 1. George Routledge & Sons, London

## Chapter 2

# Decoupling

*If it is very easy to substitute other factors for natural resources, then there is in principle no “problem”. The world can, in effect, get along without natural resources.*

– Robert Solow (1974, p. 11)

Robert Solow’s above statement was part of a wider response from neoclassical economists to the release of Donella and Dennis Meadow’s *The Limits to Growth* (1972). According to simulations created by Meadows and colleagues, indefinite future growth was unlikely, if not impossible. Rather, their model showed robust tendencies toward inevitable collapse (from both resource depletion and pollution accumulation). The *Limits* report, coupled with the onset of the 1973 oil crisis, led to heightened concern over resource depletion. Neoclassical economists, in turn, responded by insisting that input substitution, along with continued technological progress, would ensure that growth could continue indefinitely.

This indefinite growth argument rests on the assumption that economic output can become *decoupled* from resource inputs. Over the 40 years since the *Limits* debate began, resource decoupling has become a major policy objective of many governments and international agencies. For instance, the European Union insists that “breaking the linkages between economic growth and resource use” should be a key policy objective (CEC 2002). The Organisation for Economic Co-operation and Development (OECD) has placed similar emphasis on decoupling, stating that decoupling indicators are “valuable tools for determining whether countries are on track towards sustainable development” (2003, p. 13).

This chapter investigates the empirical evidence for decoupling. While the trends in the energy intensity of US real gross domestic product (GDP) seem to support the neoclassical decoupling assumption, further investigation reveals that this issue is mired by fundamental epistemological difficulties. I argue that both the evidence *for* and the concept *of* decoupling are *artifacts*. I attempt to demonstrate that the evidence for decoupling is a methodological *artifact* that results from the use of monetary value to measure the quantity of output. Furthermore, I argue that the very notion of “economic output”—and hence the notion of “decoupling”—is a conceptual artifact that results from the misapplication of linear thinking onto a complex, nonlinear system. I offer the alternative hypothesis that monetary value,

rather than representing the quantity of output, functions as a *feedback* mechanism that controls the flow of energy through the economy. I provide evidence that supports this hypothesis, and discuss the implications of this paradigm shift for a biophysical growth theory.

## 2.1 Natural Resources in Neoclassical Growth Theory

David Stern observes that “there is an inbuilt bias in mainstream production and growth theory to down-play the role of resources in the economy” (2004, p. 38). The reader may have noticed, from Chap. 1, that natural resources are completely absent from the Solow–Swan model, including its production function. Herman Daly (as previously stated) notes that “since the production function is often explained as a technical recipe, we might say that Solow’s recipe calls for making a cake with only the cook and his kitchen” (1997, p. 261). Joseph Stiglitz (1974) rectifies this shortcoming by introducing a modified production function with natural resources  $R$ :

$$Y = AL^{\beta}K^{\alpha}R^{\lambda} \quad (2.1)$$

Before proceeding to the implications of this modified production function, it is worth discussing why, in practice, it is almost never used. Neoclassical growth theory is embedded in a larger neoclassical framework in which factor incomes are assumed to be proportional to each factor’s marginal productivity. The result is that the exponents in the production function represent both a factor’s share of national income and its *relative importance to overall production*.

This condition automatically discounts the importance of natural resources, since the income of natural resource owners ( $\lambda$ ) represents an exceedingly small portion of total income. For instance, the value of total primary fossil fuel consumption constituted an average of only 3 % of US GDP over the last 60 years, meaning that when adding energy to a neoclassical production function, the energy input must be raised to the exponent 0.03, rendering it almost inconsequential.<sup>1</sup> Thus, maintaining compatibility with neoclassical distribution theory guarantees that resources will remain theoretically unimportant in neoclassical growth theory.

For the moment, however, we leave this difficulty behind and focus on Stiglitz’s modified production function (Eq. 2.1). This model presents two possible methods for decoupling output ( $Y$ ) from resource input ( $R$ ): either labor ( $L$ ) and capital ( $K$ ) inputs may be substituted for natural resources, or technological progress ( $A$ ) may shift the entire production function, allowing more output per unit of input. Since there are no inherent limits (in neoclassical theory) to either substitution or

---

<sup>1</sup> Value of fossil fuel production from Enersy Information Administratio (EIA), Annual Energy Review, Table 3.2. Value of fossil fuel imports from ibid, Table 3.7. Nominal GDP from BEA Table 1.1.5.

technological progress, there are no inherent limits to resource decoupling. This constitutes the following implicit assumption:

**Decoupling Assumption:**

The quantity of economic output can become decoupled from energy inputs.

## 2.2 Measuring Decoupling

In order to test the decoupling assumption, we must first calculate output  $Y$ . Recall that output, in the Solow–Swan model, is conceived as a *single* commodity. In the real world, many different commodities are produced, meaning some sort of aggregation is necessary. This requires that we add together the quantity ( $Q$ ) of all the commodities that are produced. Using summation notation, this becomes:

$$Y = Q_1 + Q_2 + Q_3 + \dots = \sum Q_i \quad (2.2)$$

However, this sum is meaningless without some common unit of measure. Economists typically choose unit-price ( $P$ ) for such measurement purposes. Thus, the quantity of production is now measured in terms of total value ( $Y_n$ ), which is calculated using the quantity-price product of each individual commodity and summed across the spectrum of all commodities:

$$Y_n = Q_1 P_1 + Q_2 P_2 + \dots = \sum Q_i \times P_i \quad (2.3)$$

Equation 2.3 is the formula for nominal GDP. A problem, however, is that a change in output can arise from both a change in quantity  $Q$  or a change in unit-price  $P$ . Since we are only interested in the former, we must adjust for *pure price change* (i.e., inflation) by dividing nominal GDP by some index of the average price of all commodities ( $\bar{P}$ ). This gives us Eq. 2.4—the general formula for real GDP:

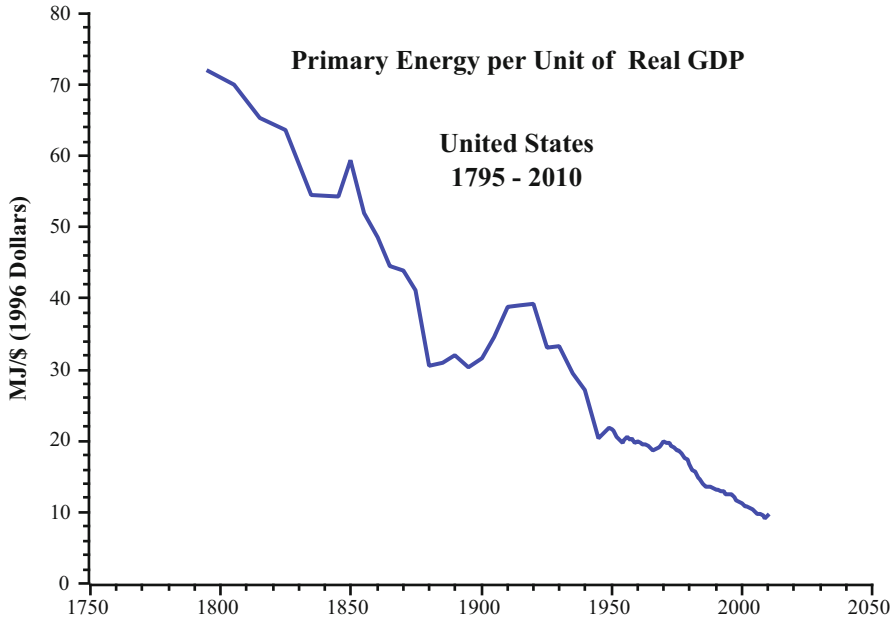
$$Y_r = \frac{1}{\bar{P}} \sum Q_i \times P_i \quad (2.4)$$

Real GDP has come to be *the* output metric used by economists. Decoupling, which is defined as an increase in output per unit of natural resource input, is typically measured in terms of the ratio of resource input  $R$ , to real GDP output  $Y_r$ :

$$\text{Resource Intensity of GDP} = \frac{R}{Y_r} \quad (2.5)$$

Since we are only concerned with energy decoupling, we replace the generalized resource term  $R$  with the more specific primary energy consumption term  $E$ , giving us the primary energy intensity of GDP:

$$\text{Primary Energy Intensity of GDP} = \frac{E}{Y_r} \quad (2.6)$$



**Fig. 2.1** Primary energy intensity of gross domestic product (GDP). (Sources: Primary energy consumption data from EIA Annual Energy Review 2011, Table 1.3 (1949–2011) & Table E1 (1795–1945). Real GDP data from HSUS Table Ca9 (1795–2000) & BEA Table 1.1.6 (2001–2010))

Figure 2.1 shows how this metric has changed over the last two centuries of the US history. While data prior to 1900 should be regarded as a rough estimate, there is clearly a long-term downward trend. It would seem, then, that decoupling is a very reasonable assumption, and we have no reason to suspect that it will not continue into the indefinite future.

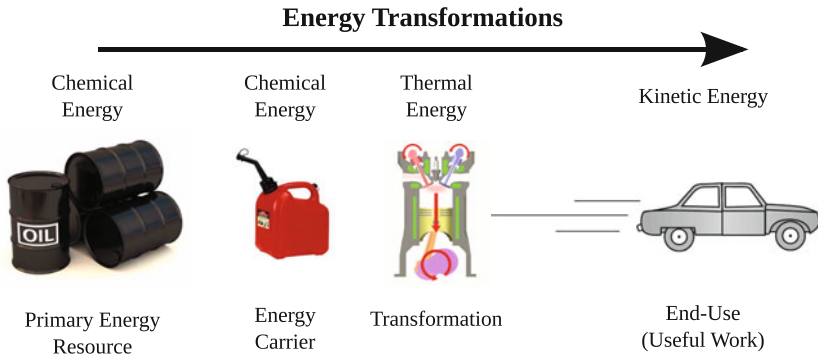
Typical investigations of decoupling go no further than this; however, there are a few complicating factors that should be investigated. The first concerns how (or “where”) we measure energy consumption. Figure 2.2 shows an example of this issue by tracing the path of energy used by a car. Energy enters the system as a primary energy resource (crude oil). This is then transformed into gasoline, then thermal energy in the car engine, and then finally the kinetic energy of the car. At each stage of the process, energy is wasted, meaning very little of the primary energy is transformed into useful work.

We can relate primary energy inputs ( $E$ ) to useful work output ( $U$ ) by means of the efficiency ( $\eta$ ) of the energy conversion process:

$$U = \eta E \quad (2.7)$$

A major difficulty arises when we move from a single process to the economy as a whole. How do we account for the efficiency of the myriad of different conversion processes that occur? Giampietro et al. (2013) argue that measuring useful work at the





**Fig. 2.2** “Where” to measure energy?

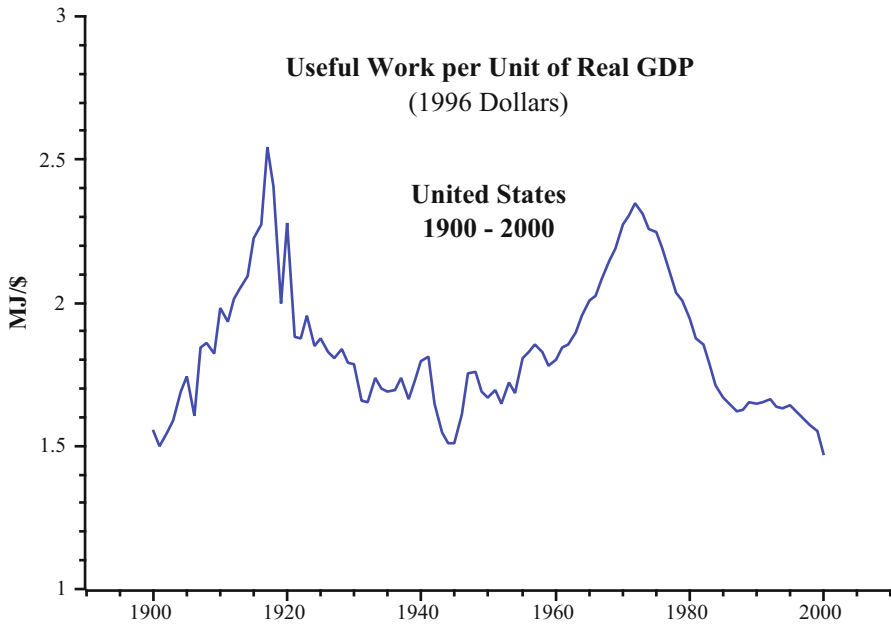
level of the entire economy is impossible. Ayres and Warr (2009), on the other hand, argue that useful work can be estimated by means of a conceptual simplification. Ayres and Warr group end-use energy into the following five categories: electricity, heat (low, mid, high), mechanical drive, light, and muscle work. They then calculate the average efficiency of the processes within each category, and then aggregate the results to get an estimate for the useful work done by the entire economy.

Using Ayres and Warr’s data, we can conceive of an alternative decoupling index which now uses useful work ( $U$ ) instead of primary energy:

$$\text{Useful Work Intensity of GDP} = \frac{U}{Y_r} \quad (2.8)$$

The time-series for the useful work intensity of real GDP (Fig. 2.3) looks nothing like the primary energy time-series (Fig. 2.1). While there is significant change in the useful work intensity of GDP, there is little evidence of long-term decoupling. This important result indicates that the majority of historical *primary* energy decoupling has been the result of efficiency increases in energy conversion processes.

This increase in efficiency is what neoclassical economists would call *embodied* technological progress, as opposed to the *disembodied* form ( $A$ ) used in the Solow–Swan model (Sakellaris and Wilson 2004). Most importantly, this embodied form of technological progress has *inherent limits* stipulated by the laws of thermodynamics. While 100 % efficiency is the ultimate limit (forbidden by the second law of thermodynamics), the maximum theoretical efficiency for many energy conversion processes (such as heat engines) is much less. For instance the Carnot Limit for the internal combustion engine is a mere 37 %, with current car engines having efficiencies of about 20 % (Chandler 2010). Thus, a future doubling of engine efficiency is out of the question. It appears that primary energy decoupling has mainly been the result of increases in energy conversion efficiency. Since there are strict limits on the efficiency of such conversion (for example, coal conversion to electricity appears to be nearly unchanged at 40 % for decades), we must concede that decoupling cannot continue indefinitely.



**Fig. 2.3** Useful work intensity of GDP. (Sources: Useful work data from Benjamin Warr REXS Database. Real GDP data from HSUS Table Ca9)

### 2.3 Problems with Real GDP

Before we continue with the investigation of decoupling, it is important to discuss some of the difficulties with using monetary value to estimate an index of output. There are three basic problems:

1. Changes in relative prices mean that the basic unit of measure (price) is not well defined.
2. Measures of quality change contain inherently subjective elements. Moreover, current quality change methodology may be circularly dependent on neoclassical theory.
3. In an economy dominated by services, it is not clear that an “output” can be objectively defined.

We begin with the problem of changes in relative price. The transmutation from the *value* of output (nominal GDP) to the *quantity* of output (real GDP) relies on the existence of an index capable of accurately and *objectively* quantifying the average change in prices. But unless price changes are completely homogeneous, the actual calculation of such an index is intrinsically subjective.

One might think that an average price is unambiguous. One simply adds the prices of all commodities together and divides by the number of commodities. But this is an *unweighted* average. If we were to compute the average price this way, we effectively

**Table 2.1** Output using explicit average price

Year	Quantity A	Quantity B	Price A (\$)	Price B (\$)	Price index	Output (\$)
1	100	500	20	10	1.0	7000
2	200	500	40	20	2.0	9000
3	200	500	80	40	4.0	9000
4	200	500	160	20	3.7	11,308

state that the change in price of a single unique commodity (i.e., a rare piece of art) is as important as the same change in price of a *billion* uniform commodities (i.e., pencils). What we want is a weighted average that reflects the fact that some prices are more important than others. But herein lies the crux: there are numerous ways that such an average might be weighted.

In order to demonstrate this issue, we must return to the theoretical underpinnings of real GDP. Equation 2.4 showed (in abstract form) the methodology used to calculate real GDP. To make things more concrete, let us apply this formula to a simple economy with two commodities (*A* & *B*). Real output ( $Y_r$ ) becomes the quantity-price product of these two commodities, divided by an index of their average price:

$$Y_r = \frac{Q_A P_A + Q_B P_B}{\bar{P}} \quad (2.9)$$

Table 2.1 shows sample calculations for output ( $Y_r$ ), using imagined numbers. Here we use an *explicit* average price, calculated by weighting the price of a commodity by its respective quantity in the first year. Notice that we have a change in the quantity of commodity *A* only between year 1 and 2. During all other years, the quantity of each commodity remains unchanged. Additionally, price change is homogeneous for the first three years: all prices double annually. However, in year 4 we introduce a divergent price change: the price of *A* doubles while the price of *B* is halved. The result of this divergent price change is an *increase* in the measure of output without *any* underlying change in the quantities of either commodity.

Alternatively, we can calculate the average price *implicitly* (which is how it is most often calculated). To do this, we first calculate the nominal value of output, and then divide this by the value of output with constant prices. But, in order to do this, we must choose a base year from which to hold prices constant. For instance, the formula for the average price with a year 1 base would be:

$$\bar{P}_{yr1} = \frac{Q_A P_A + Q_B P_B}{Q_A(\$20) + Q_B(\$10)} \quad (2.10)$$

The decision about which base year to use is completely subjective, yet it is nontrivial. It can have major effects on our measure of output. For instance, the results for both a year 1 base ( $Y_{r1}$ ) and a year 4 base ( $Y_{r4}$ ) are shown in Table 2.2. While the erroneous output change between year 3 and year 4 has been corrected, we now have a time-series *discrepancy* associated with the use of different base years. There are no objective grounds for deciding which series is correct. Our measure of output is simply ambiguous.

**Table 2.2 Output using implicit average price**

Year	Nominal value (\$)	Price index ( $Yr_1$ )	Price index ( $Yr_4$ )	Output ( $Yr_1$ ) (\$)	Output( $Yr_4$ ) (\$)
1	7000	1.0	1.0	7000	7000
2	18,000	2.0	1.6	9000	11,340
3	36,000	4.0	3.2	9000	11,340
4	42,000	4.7	3.7	9000	11,340

As demonstrated by these examples, an average price (and any measure of output derived from it) is uniquely defined *only* when price change is *homogeneous*. In the case where the relative price change between commodities is *heterogeneous*, any calculation of the average price will be but one of many that are possible. Additionally, the degree to which such calculations may diverge is a function of the degree to which relative prices diverge.

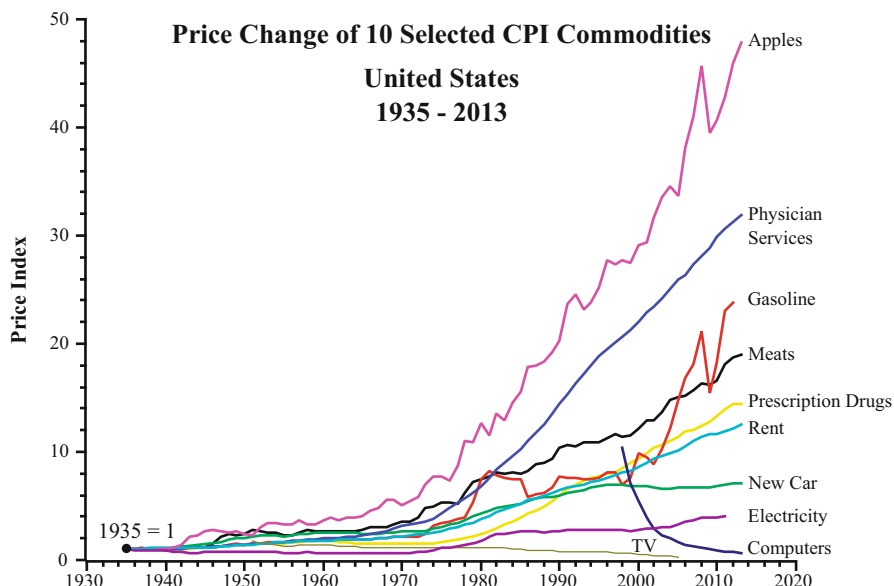
The way to understand this phenomenon is to realize that prices constitute our basic unit of measurement. By definition, the most important attribute of a unit is its *uniformity*. When prices change uniformly over time there is no problem—we simply dilate (or compress) our unit accordingly. However, if price change is *not* uniform for all commodities, we have a fundamental problem: our unit is *no longer well-defined*. In principle, there is no way to escape this dilemma. Without a well-defined unit, objective measurement is impossible. In such a case, many different values of real GDP can be defined, each backed by a different method of calculating the average price. At most, one can subjectively argue for the value of real GDP that is “best”. However, this subjectivity means that we are no longer measuring what output *is*, but rather, what we think it *should be*.

Thus, it is important to investigate price change in the real world. If price change is relatively homogeneous, the preceding theoretical hurdle can be dismissed as unimportant. On the other hand, if price change is *divergent*, then we must seriously question the objectivity of real GDP. Figure 2.4 shows historical US price changes for a select group of consumer price index (CPI) commodities. Although this is by no means a representative sample, it does serve to illustrate the enormous degree to which actual prices have diverged over the twentieth century.

To give a sense of the scale of this divergence and to quantify the ambiguity in our unit, we can calculate the relative standard deviation of these prices.<sup>2</sup> Initially zero in 1935 (the base year), by 2010, the relative standard deviation in the prices of these ten commodities was 91 %. This indicates that over the course of 75 years, the normal *divergence* from the average price was almost as large as the average itself.

It is important to remember that this is not a measurement *error*—the result of the imprecision of a measuring device. In order to state such an error, one must have a stable unit. For instance, we can state that a thermometer measures the temperature as 20.2 °C with a measurement error of  $\pm 0.5$  °C. But notice that the error estimate

<sup>2</sup> Relative standard deviation is defined as the standard deviation divided by the data mean and multiplied by 100 %.



**Fig. 2.4** Diverging prices. (Source: BLS Consumer Price Index, all urban consumers. Commodities added after 1935 are indexed to the unweighted average price for the year in which they are introduced)

is expressed in the same units as the measurement itself. Thus, to even *state* this error requires that we agree on the definition of our basic unit (the degree Celsius is precisely defined as 1/100th of the temperature range between the freezing and boiling point of water at sea level). Unlike measurement error, what Figure 2.4 shows is a *fundamental and irreducible* uncertainty in our *basic unit*.

This can be put in perspective by comparing the problem to units of length. Historically, people measured small lengths in units of body parts (i.e., actual “feet” and “hands”). Imagine a similar unit called the “man”, defined simply as the height of the man doing the measuring. This unit is not very well defined, since height varies considerably among men. However, in contrast to the 90 % deviation in relative prices calculated above, the relative standard deviation in the height of adult males is only about 4 % (Smith et al. 2000). This indicates that, as inaccurate as a unit of length called a “man” would be, it is still *20 times more accurate* than the unit on which real GDP is based. One is tempted to imagine how many bridges would remain standing if engineers based their calculations on such a poorly defined unit.

Since the choice of base year effectively “defines” the unit of measure, rebasing (changing the base year) can lead to spectacular changes in real GDP. For instance, Nigeria recently changed from a 1990 to a 2010 base year. The result was that its GDP nearly doubled (Blas and Wallis 2014). A similar doubling of Ghana’s GDP occurred in 2010 when it changed its base year from 1993 to 2006. Base year revisions in Botswana, Kenya, Tanzania, and Zambia have also led to large changes in GDP (Jerven 2012; 2014).

In response to this difficulty, the US Bureau of Economic Analysis (BEA) has changed how it calculates real GDP. It has moved from the base year method to a “chain-weighted” approach that uses an average of multiple, moving base years (Steindel 1995). The hope is that this will “correct” the errors introduced by divergent prices. My contention, however, is that this misses the point. From an epistemological standpoint, the validity of a unit of measurement hinges on it being well-defined over time and space. Since prices (our basic unit) change in nonuniform ways, they fail this basic condition. The chain-weighting approach attempts to “solve” a problem that is unsolvable.

The BEA makes the claim that a chain-weighted average price is “better” than a non-chain-weighted one. However, such a claim is untestable: because we are dealing with the ambiguity of the basic unit, there are no objective criteria on which to determine what is “better” or “worse”. The only objective way to determine if a scientific measurement is “better” is if it is more accurate (i.e., has smaller error). But as discussed above, to determine such an error paradoxically requires a well-defined unit of measurement. The basic instability in the unit of price cannot be altered by our methodological choice of how to compute the average price; instead it is an inalterable function of *history*.

A further problem, when moving from neoclassical theory to the real world, is that the former posits a single, *unchanging* commodity. In reality, the *qualities* of a commodity can change immensely with time (think of the difference between a 1980s computer and a modern laptop). In price index methodology, a *quality* change is taken to be the same thing as a *quantity* change. Jonathan Nitzan explains:

[S]uppose Ford Motors produced 100,000 Mustang cars at a unit price of \$ 10,000 in 1975 and manufactured 150,000 units at a price of 14,000 per car in 1985. If we can presume that the Mustang of 1975 was identical to the one produced in 1985, we can, without ever defining what a Mustang is, conclude that there was a 50 percent increase in quantity and a 40 percent rise in price. On the other hand, if we acknowledge that the two models are different, such a direct comparison has little meaning and we must now both define the ‘commodity’ and describe how it changes over time.

The two Mustang models may vary in aspects of production—such as the technology with which they were manufactured, the labour involved in their assembly, and their material composition. They could also vary in their so-called ‘consumption attributes’—such as weight, size, power, shape, speed, comfort, colour, fuel efficiency, noise and chemical pollution. Under such circumstances, we must somehow denominate all such ‘quality’ differences in universal, quantitative terms and adjust our computations accordingly.

For instance, if because of such changes, a 1985 model contained twice as much ‘automobile quality’ as the 1975 model, we would have a 200 percent rise in quantity produced and a 30 percent decrease—not increase—in unit price! On the other hand, if quality was found to be 50 percent lower in the 1985 model than in the 1975 one, we would end up with a 180 percent rise in price and a 25 percent reduction in quantity! (1992, p. 156)

Thus, the objective quantification of output hinges on the objective quantification of qualitative change. Note that the scale of such adjustments is not trivial. For instance, Fig. 2.4 shows massive deflation in the price of computers. While actual unit prices have declined slightly over the last 30 years, the majority of this deflation is a result of the Bureau of Labor Statistic (BLS)’s quality change adjustment. The BLS looks at variety of computer attributes (processor speed, hard drive capacity, ram, operating

system, etc.) and attempts to measure how these attributes have improved over time. It then combines the changes in each attribute (which involves deciding the relative importance of each attribute) to arrive at an index for the overall change in “quality” of the computer. This “quality” then has an inverse effect on the price index: if the “quality” of a computer doubles, the price index is halved.

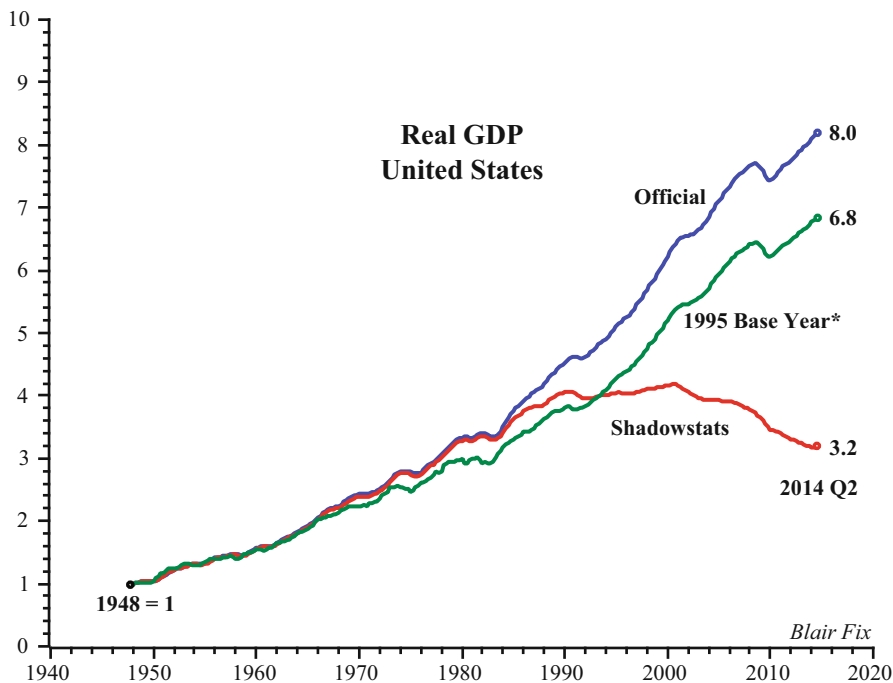
While quality adjustment is an essential element in the construction of a price index, there are numerous conceptual difficulties associated with such adjustments (UN 1977, Nitzan 1992). For instance, in order to quantify the changing nature of a commodity, one must first subdivide the commodity into relevant “attributes”. But how do we *objectively* decide those attributes that are relevant and those that are irrelevant? Furthermore, once we have reduced a commodity to its constituent attributes, how do we decide their relative importance? The most popular method is called “hedonic quality adjustment”. The BLS (2010) summarizes the process as follows:

In price index methodology, hedonic quality adjustment has come to mean the practice of decomposing an item into its constituent characteristics, obtaining estimates of *the value of the utility* derived from each characteristic, and using those value estimates to adjust prices when the quality of a good changes. (emphasis added)

All quantitative comparisons require a unit of measurement. Here we see that the BLS is attempting to measure the attributes of a commodity in units of *utility*. This is problematic for two reasons. First, utility is a hypothetical psychic flux that cannot be directly measured, making it unsuitable as a basic unit. Because utility cannot be observed, Nitzan (1992) argues that hedonic quality measurements are unfalsifiable. Secondly, utility is a fundamental tenet of neoclassical theory. By locating “quality” in a commodity’s ability to give utility, statistical agencies are explicitly adopting a neoclassical theory of value (i.e., that value comes from utility). Nitzan writes: “Both the idea that quality can be measured (objectively or not) and the methods developed for that purpose are closely tied with the neoclassical paradigm. The evidence supporting these conclusions seems overwhelming” (1992, pp. 175–176).

This possible circularity between neoclassical theory and the estimation of real GDP represents a fundamental problem for the empirical researcher. It is a cardinal sin to test a theory using data that is circularly dependent on the theory being tested. As such, if we want to conduct a truly independent test of neoclassical growth theory, we may need to look for alternative metrics. Furthermore, we should not discount the fact that the national accounts are created by and for governments. Since governments are under immense pressure to show that the economy is growing, we should not be so naive as to treat official real GDP estimates as a “disinterested” measurement.

The last problem with real GDP is that while the notion of a “quantity of output” is intuitive when applied to the production of *goods*, it becomes less intuitive (even nonsensical) when applied to certain *services*. For instance, how do we objectively define the output quantity of education, healthcare, the performing arts, or of finance? Imagine an economy that initially consists of only farmers, but gradually changes its composition to consist of only financiers. In dollar terms, the economy will grow, because financiers earn more money than farmers (in economic jargon, they “add



**Fig. 2.5** Divergence between official and alternative real GDP estimates. (Source: Official real GDP from BEA Table 1.1.6. “Shadowstats” GDP from John Williams’ Shadow Government Statistics ([shadowstats.com](http://shadowstats.com)). ‘Vintage’ 1995 base year GDP from Federal Reserve Bank of Philadelphia (ROUTPUT95Q1). Note that this data ends in 1995, but I have projected it forward (for comparison purposes) using official growth rates)

more value”). Yet in physical terms, the economy will shrink, because financiers do not produce anything tangible.

The enormous structural shift toward services means that this thought experiment is now playing out in reality: services have been a major source of US GDP growth in the last few decades, yet they have no clear physical dimension. Statisticians avoid the problem of measuring service “output” by simply *defining* the output of such services as equivalent to their market value, and then deflating this value using a price index from another sector that is more amenable to measurement (Sherwood 1994; Griliches 1992). Putting aside all of the measurement problems discussed above, the fact that the majority of people in the USA are now employed in services means that the very notion of a “quantity of output” is becoming increasingly fuzzy. If we cannot objectively define what output *is*, we cannot objectively measure it.

At this point, if one accepts my argument that the estimation of real GDP is inherently subjective, one might still object that the scale of the problem is not that large. Indeed, when stated only in theoretical form, it is difficult to grasp the degree to which different estimates of real GDP might diverge (when different methodological



choices are used). This difficulty can be remedied by contrasting official and unofficial estimates of real GDP. Figure 2.5 plots the official real GDP estimate against estimates from John Williams' Shadow Government Statistics website and a "vintage" estimate using a 1995 base year (curiously the year that the BEA abandoned the fixed base year method). The discrepancy between upper and lower estimates is a whopping 250 %. Again, it must be stressed that there are no objective criteria to choose between these three estimates.

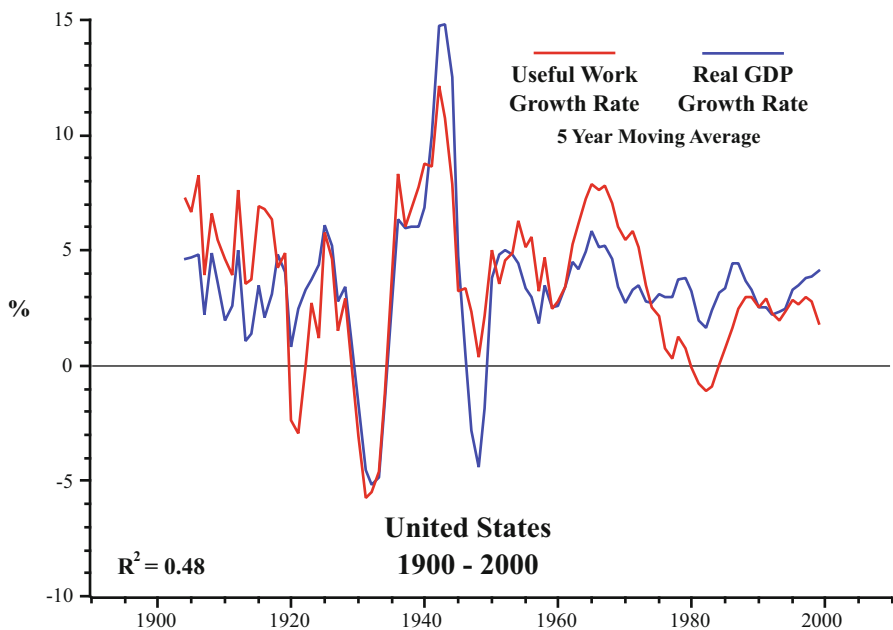
The basic theoretical edifice of "decoupling" rests upon both a well-defined resource input, and a well-defined output. My contention, however, is that the lack of a stable unit makes it *impossible* to objectively define a quantity of output. This puts the whole notion of decoupling into a measurement limbo.

## 2.4 Decoupling: Fact or Artifact?

For the remainder of the chapter, I investigate the hypothesis that the evidence for decoupling is a methodological *artifact* that arises from the use of monetary value to measure output. As demonstrated below, when the price of energy is used to deflate nominal GDP (rather than the GDP deflator), evidence for decoupling almost entirely disappears. I hypothesize that monetary value, rather than represent the quantity of output, functions as a feedback device for controlling the flow of resources. Further investigation suggests that this feedback is not random; rather, it is fundamentally related to the biophysical labor productivity of the mining sector.

To frame this discussion, we return to evidence for the decoupling of real GDP from useful work. However, rather than look at the *ratio* of useful work to real GDP (as we did in Fig. 2.3), we now look at the respective *growth rates* of each series (i.e., we move from absolute units to percentage rate of change). Figure 2.6 plots the historical growth rates of useful work and real GDP over the last century. When plotted this way, evidence for decoupling will appear as a divergence in the growth rates of the two series, with real GDP growing more quickly than useful work. Note that for most of the period shown, the two series are very similar, meaning there is little evidence of decoupling. However, after 1970, useful work and real GDP growth rates diverge significantly, with real GDP continuing to grow at about 3 % per year, while useful work growth slows significantly. What is behind this decoupling? Does it represent some new form of technological progress, some sort of capital substitution, or the take-off of the "knowledge economy"?

Warr and Ayres (2012) attempt to explain this decoupling by looking at the growth of information and communications technology capital. I pursue a different explanation. Instead, I hypothesize that this divergence is an *artifact* of our measure of output, real GDP. In order to test this hypothesis, we must first adopt a different method for deflating nominal GDP. Recall that the transmutation from nominal to real GDP ( $Y_n \Rightarrow Y_r$ ) is accomplished by means of an average price index ( $\bar{P}$ ) that



**Fig. 2.6** Annual growth rates of real GDP and useful work. (Sources: see Fig. 2.3)

adjusts for inflation:

$$Y_r = \frac{Y_n}{\bar{P}} \quad (2.11)$$

Nearly all of the theoretical difficulties associated with real GDP reside in the construction of the average price index. A simple way around these problems is to avoid the average price concept entirely, and instead deflate nominal GDP by the price of a *single, unchanging* commodity. By doing so, however, we are no longer constructing a quantity measure of output; rather, we are constructing what I call a *specific purchasing power (SPP)* index<sup>3</sup>. The specific purchasing power of commodity  $x$  is defined as nominal GDP divided by the price of  $x$ :

$$SPP_x = \frac{Y_n}{P_x} \quad (2.12)$$

The specific purchasing power of commodity  $x$  measures the ability of a nation to *finance* the consumption of this commodity. It is a symbolic measure in the sense that it denotes neither the amount of commodity  $x$  that is actually consumed nor the

<sup>3</sup> I use the word *specific* to disambiguate from the more common notion of purchasing power, which typically refers to the ability to consume a basket of goods or services, rather than a single good.

amount that could conceivably be consumed.<sup>4</sup> However, since the ability to finance consumption is a prerequisite to *actual* consumption, this metric should be related to biophysical flows.

In order to avoid the problem of measuring changes in quality, we want to choose a commodity that remains uniform across time. Energy resources are almost completely unique in this regard. While primary energy sources come in many qualitatively different forms, the science of energetics allows us to compare all of them using a single, well-defined unit (the Joule). In this section, I use energy-specific purchasing power to avoid average price problems.

We can use energy-specific purchasing power to see if the post-1970 decoupling shown in Fig. 2.6 is a methodological artifact. However, by abandoning real GDP in favor of specific purchasing power, we are no longer investigating the decoupling of output from resource input; rather, we are looking for evidence supporting the following monetary feedback hypothesis:

**Monetary Feedback Hypothesis:**

Monetary value represents a feedback mechanisms for controlling resource flows. As such, the growth of energy inputs should be tightly coupled to the economy's ability to finance their consumption.

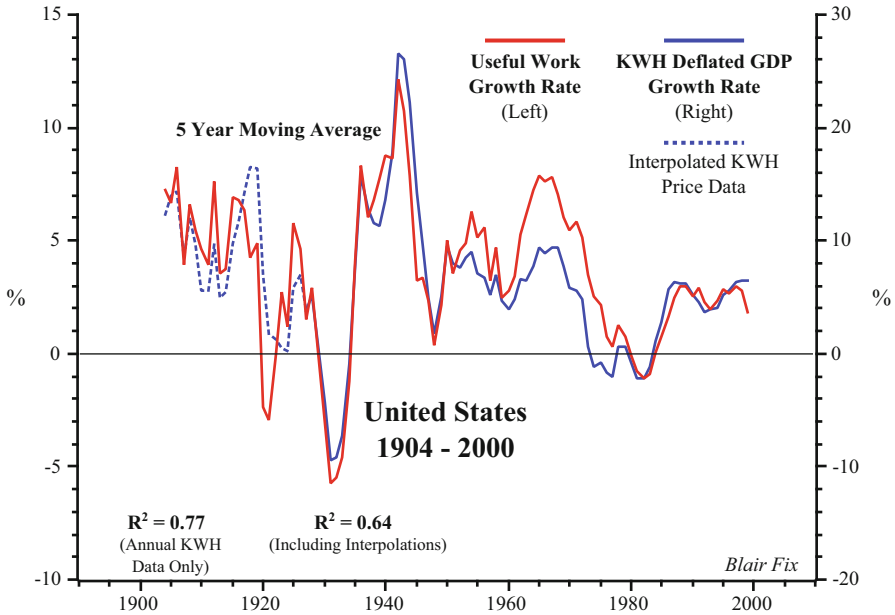
We can test this hypothesis by investigating changes in the specific purchasing power of useful work (which is equal to nominal GDP deflated by an index of the price of useful work). An immediate problem is that most forms of useful work have no price. We usually pay for energy in its primary form; once it has been transformed into an end-use application, we almost never pay directly for this work. For instance, we purchase the gasoline used by a car, but *never* the end-use kinetic energy of the car. The only Ayres–Warr category of useful work with a well-defined market price is *electricity*. Deflating nominal GDP by the price of a kilowatt-hour (kWh) of electricity gives *kWh-specific purchasing power*:

$$SPP_{kWh} = \frac{Y_n}{P_{kWh}} \quad (2.13)$$

My contention is that kWh-specific purchasing power represents a rough proxy for the ability of the economy to purchase useful work. We are not used to thinking in these terms, so a comparison to the household level is helpful. One's household income determines what one can consume. If we compare one's income to the price of a pencil, this gives an indication of how many pencils one could consume. By extension, if we compare the value of the entire economy to the price of a pencil, we get an indication of how many pencils the economy could consume. Likewise, by comparing the value of the entire economy to the price of useful work (proxied by

---

<sup>4</sup> While an individual could conceivably spend his entire income on a single commodity, it is impossible (and absurd) for a nation to do so. For instance, if the USA spent its entire nominal GDP on fossil fuel energy, its energy consumption would be greater than that for the entire world, and no other goods or services would be generated.

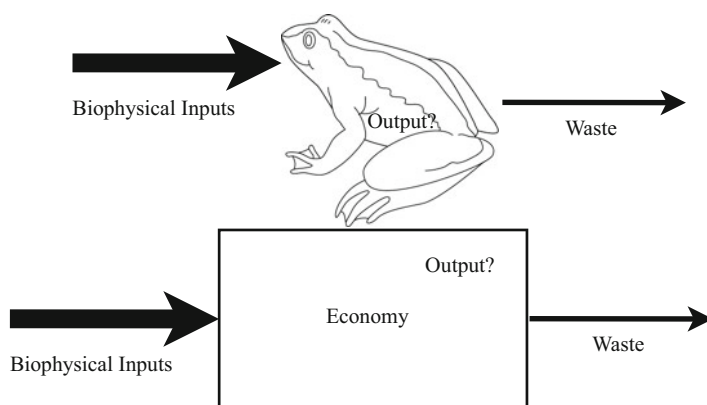


**Fig. 2.7** Growth rates of kWh-specific purchasing power and useful work. kWh-specific purchasing power represents a rough estimate of the ability of the economy to finance the consumption of useful work. (Sources: Nominal GDP from HSUS Table Ca10 (1900–1928) and BEA Table 1.1.5 (1929–2000). Price of electricity (1904–2000) from HSUS Table Db234 (average price, all services). Electricity prices for 2001–2011 calculated by weighted average over three end-use sectors: commercial, industrial, and residential. Sectoral prices from EIA Annual Energy Review 2011, Table 8.10; consumption from Table 8.9)

the price of electricity) we get an indication of the economy’s ability to finance the consumption of useful work. If the monetary feedback hypothesis is true, the growth rate of kWh-specific purchasing power should be highly coupled to the growth rate of useful work.

When we compare the historical growth rates of kWh-specific purchasing power to those of useful work (Fig. 2.7), an interesting thing happens: the post-1970 decoupling of real GDP from useful work (shown in Fig. 2.6) almost entirely *disappears*. Throughout the entire twentieth century, the growth in the US economy’s ability to purchase electricity remained tightly correlated with the growth of useful work. This evidence is consistent with the monetary feedback hypothesis, and it supports the hypothesis that the evidence for decoupling is a measurement *artifact*.

I should be clear that my argument is not that statistical agencies have somehow made a “mistake” in their calculation of output. To the contrary, I hypothesize that the notion of “output” (and therefore, “decoupling”) is a *conceptual artifact* that results from the misapplication of linear thinking to a nonlinear system. If we think in biophysical terms, the economy is a complex, nonequilibrium system that uses



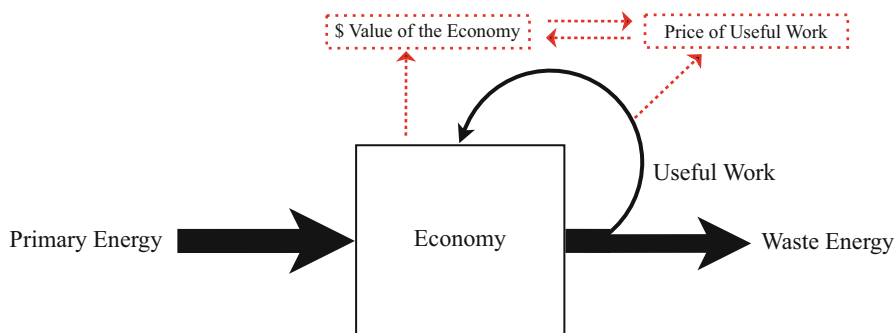
**Fig. 2.8** The biophysical flows through a frog and the economy. Where is “output”?

biophysical flows to sustain itself. The only linear output of such a system is its *waste*.

This can be made more clear if we compare an economy to a living organism (in this case, a frog). Both the frog and an economy function by exploiting a flow of biophysical inputs and exuding a flow of biophysical waste outputs (Fig. 2.8). This material flow sustains the internal processes of each system. While the frog has many internal subsystems/organs that might be said to produce an “output” (i.e., the heart has a blood flow output, the lungs an output of oxygen, and the endocrine system an output of hormones), all such outputs are destined to become inputs to other processes. Thus, the internal workings of an organism are intrinsically circular—the frog (as a whole) has no “output” other than its waste.

When we think in these terms, the economy, like the frog, has no output; rather, it has a resource *throughput*. Our mistake comes when we label certain internal processes as “output”: This gives the illusion of linearity where none actually exists. All of the outputs of the myriad of internal processes within the economy are destined to become inputs to other processes. Thus, the internal workings of the economy are inherently circular, meaning the notion of a linear output is difficult to justify.

I argue that the notion of “output” (at the level of the entire economy) is a conceptual artifact that arises from the focus on monetary value. That is, we conflate a *sale* (a monetized exchange) with the creation of an *output*. Note that a sale is inherently linear: Money always flows from buyer to seller. Thus by aggregating sales (and calling this output), we create the illusion that the economy is a linear process. However, if we drop the assumption that a sale represents an output, the illusion of linearity disappears: all internal processes become circular and the very notion of output (and hence, decoupling) becomes untenable. At the level of the entire economy, the only linear flow is the stream of biophysical throughput, which ends in the output of waste.



**Fig. 2.9** Monetary value as a feedback device for “controlling” biophysical flows

Rather than treat monetary value as an output, I offer the alternative hypothesis that monetary value functions as a feedback device for controlling the flow of biophysical throughput (Meadows 2012). We can frame this paradigm shift by asking the following question: how does the economy “know” to consume more resources? In the animal kingdom, the stimulus to consume resources comes from sensory feedback: animals “know” to consume resources because they “feel” hungry. What is the corollary of this sensory feedback in the economic system? My hypothesis is that monetary value functions as such a feedback mechanism, stimulating or stifling the flow of resources. Figure 2.9 shows a schematic of this process when applied to the flow of useful work.

The exploitation of a resource does not intrinsically require monetary value. However, if an economy becomes monetized (meaning certain human interactions require the exchange of money), then resource exploitation is suddenly restricted by the stock of money (since resources must be “paid for”). The pool of monetary value by which we finance the consumption of resources is perpetually renewed by a process that economists call “adding value”. For any given internal process, the sale of the final product always has a greater value than the sum of the inputs. The sum of all such added value is equal to nominal GDP—the value of the entire economy.

As this pool of added value expands relative to the price of a resource, we feel “wealthier”. Resources are “cheap” so we consume more of them. If the opposite is true—the value of the economy contracts relative to the price of a resource—we feel “poorer”. Resources are “expensive” so we consume fewer of them. Thus, I argue that prices constitute a feedback system that regulates the flow of resources through the economy. This feedback system functions so long as humans agree not to consume resources unless they can be “paid for” (i.e., we agree not to *steal*). The long-term coupling between the growth rate of useful work and kWh-specific purchasing power lends credence to this view.

By thinking in this way, however, we place a heavy emphasis on the price of energy (the price of electricity in this case). Thus, we must ask—where does the price of energy come from? It is rather disconcerting to think that random market fluctuations might cause a change in the price of energy that somehow leads to a

change in the entire economy's ability to consume useful work. This would lead us straight back to the neoclassical view that the market is the ultimate arbiter of the economy. The task of biophysical economics should be to show that energy prices are, in fact, not random at all. Instead, they are a reflection of a broader biophysical reality.

Interesting work on the topic of energy prices has been done by King and Hall (2011), who link historical prices of oil and gas to energy return on investment (EROI). Giampietro et al. (2012) stress that the biophysical labor productivity of the energy sector is equally as important as EROI in indicating the quality of the energy production process. Building on Giampietro et al., we can define the biophysical labor productivity of the US mining sector ( $\rho_M$ ) as the energy content of domestic fossil fuel production ( $E_{FF}$ ) divided by the number of workers in the mining sector ( $L_M$ ):

$$\rho_M = \frac{E_{FF}}{L_M} \quad (2.14)$$

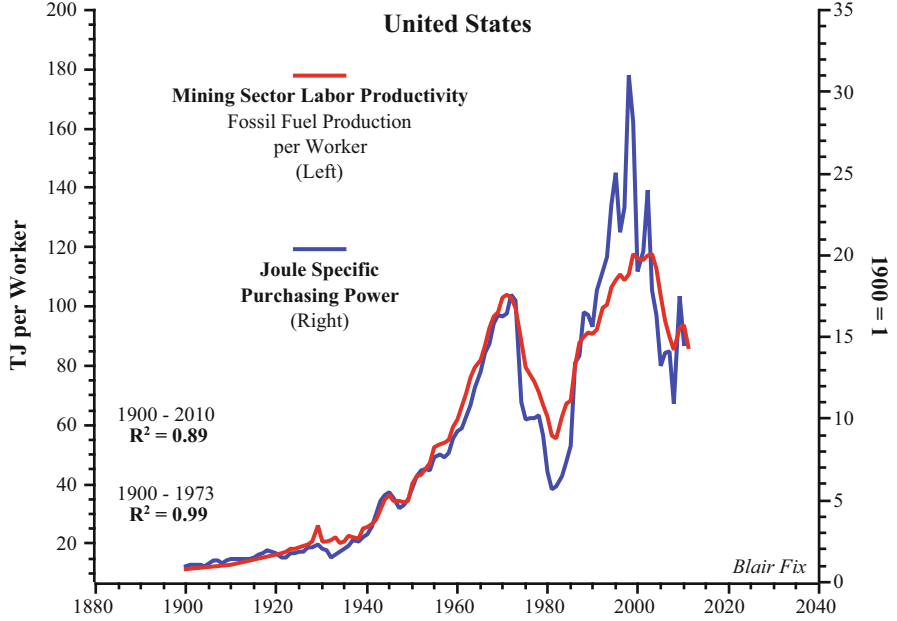
As King and Hall have done with EROI, it seems quite reasonable to hypothesize that the price of energy could be somehow linked to mining productivity. Before proceeding, it is helpful to first reflect on what the price of a commodity really means. Most economists would agree that the nominal price of a commodity is meaningless, in and of itself. A price only gains meaning when placed in comparison to the price of *other* commodities. One way of making such a comparison is to relate the price of a *single* commodity to the average price of *all* commodities. This gives us the “real” price of a commodity. However, as discussed in Sect. 2.3, there are numerous problems with average price indexes, so I avoid this approach. A different (very unorthodox) way of contextualizing the price of a single commodity is to compare it to the *total value* of all (new) commodities. Unlike the average price, the total value of all new commodities (nominal GDP) is uniquely defined at a single point in time (sampling errors aside).

Based on this unorthodox approach, we can contextualize the price of energy by comparing it to the value of all new commodities—nominal GDP. Since we are now concerned with the output of the mining sector (fossil fuels), we are interested in the nominal price of one domestically produced Joule of fossil fuel energy ( $P_J$ ). We calculate this price by dividing the total value of fossil fuel production by the energy content of this fuel. Using the notation developed previously, *joule-specific purchasing power* is defined as the ratio of nominal GDP ( $Y_n$ ) to the average price of one Joule of fossil fuel:

$$SPP_J = \frac{Y_n}{P_J} \quad (2.15)$$

Figure 2.10 plots joule-specific purchasing power against the biophysical labor productivity of the mining sector. The two series show a striking degree of correlation. Given this high correlation, it seems reasonable to write the following proportionality statement:

$$SPP_J \propto \rho_M \quad (2.16)$$



**Fig. 2.10** Joule specific purchasing power (the ability of the economy to purchase fossil fuels) and mining sector biophysical labor productivity

Since what we are interested in is the price of fossil fuel, we use the definition of joule-specific purchasing power (Eq. 2.15) to arrive at Eq. 2.17:

$$\frac{Y_n}{P_J} \propto \rho_M \quad (2.17)$$

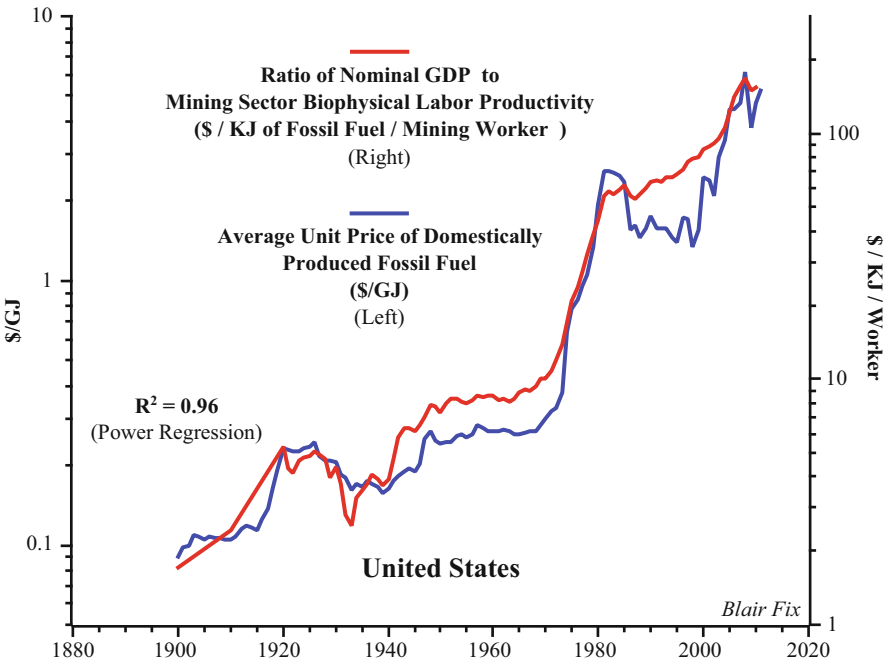
By rearranging for fossil fuel price, we get Eq. 2.18:

$$P_J \propto \frac{Y_n}{\rho_M} \quad (2.18)$$

This indicates that the nominal price of fossil fuel is a simple function of two variables: nominal GDP and the biophysical productivity of the mining sector. Figure 2.11 plots the left and right sides of Eq. 2.18 as separate time series. The results are robust: the price of fossil fuel can be almost completely accounted for by the ratio of nominal GDP to mining sector productivity (implying that biophysical considerations have a great power in our economy).

Thus, we can conclude that the price of energy is not arbitrary. Rather, when contextualized against the value of the entire economy, the price of energy seems to be fundamentally determined by the biophysical productivity of the mining sector. This is an important result, because it allows us to ultimately ground monetary feedback in what is arguably the most important process of an industrial economy: the exploitation of fossil fuels.





**Fig. 2.11** Biophysical underpinnings of the price of fossil fuels. The relation between the nominal price of fossil fuels and the nominal value of the economy is arbitrated by the biophysical labor productivity of the mining sector. (Sources for both figures: Fossil fuel production data from HSUS Table Db155 (1920–1948) & EIA Annual Energy Review 2011, Table Mining sector employment from HSUS Ba819 (1900–1910), HSUS Ba841(1920–1957), & BLS CEU1021000001 (1958–2011). Value of fossil fuel production (1949–2010) from EIA Annual Energy Review Table 3.2. Natural gas price (1930–1949) from EIA Historical Natural Gas Annual 1930 Through 2000, Table 1. Crude oil (1920–1949) from BP Statistical Review of Energy 2011. Anthracite (1920–1949) from HSUS Tables Cc238–240)

To summarize our results, we find that the evidence for decoupling almost completely disappears when nominal GDP is deflated by the price of electricity, rather than by the GDP deflator. Evidence also suggests that the relation between nominal GDP and the price of fossil fuel is arbitrated by the biophysical productivity of the mining sector. I have hypothesized that this implies that evidence for decoupling is a methodological artifact—a result of the decision to measure output in terms of monetary value. The evidence presented here supports the alternative hypothesis that monetary value functions as a feedback device for controlling biophysical throughput.

## 2.5 Decoupling in Theory and Reality: What Goes Wrong

The neoclassical argument for decoupling is simple and intuitive: either through technological progress or input substitution, the economy can become more *efficient* at transforming natural resources into final output. However, when we attempt to apply this theoretical argument to the real world by *measuring* the decoupling of output from input, we find that the concept of decoupling is plagued by fundamental epistemological problems.

Decoupling metrics are a subset of a larger class of output/input metrics in which the efficiency of a system is defined in terms of its output per unit of input. In my view, all sound efficiency metrics must satisfy three basic requirements: (1) flow consistency, (2) boundary consistency, and (3) unit consistency.

*Flow consistency* means that we should track the flow of the *same* substance through the system. For instance, when measuring the efficiency of an internal combustion engine, we continuously follow the flow of *energy* through the system. Thus, engine efficiency is defined as the useful energy output per unit of total energy input. If we do not maintain flow consistency, then the notion of efficiency becomes ambiguous. For instance, the efficiency of a computer is not easily defined. If we track energy inputs, we run into the problem that the energetic output of a computer (the work done on the electrons in its circuits) is of little interest to computer users. Instead, computer users care about outputs of processor speed, information storage capacity, and a host of other qualities. Thus, to calculate the “efficiency” of a computer we must break flow consistency and measure output on a different basis than input. By doing so, we forgo the possibility of a single metric for the efficiency of a computer; instead, we get a different efficiency metric for each relevant output that is chosen.

Moving on to *boundary consistency*, Giampietro et al. (2013) note that well-defined system boundaries are a prerequisite for any quantitative analysis. A boundary definition allows us to differentiate between what is “inside” and what is “outside” the system. The consistency of boundary definitions is essential for the calculation of an efficiency metric. For instance, when calculating the efficiency of a car, one must stipulate whether the system includes the *entire* car, or only the passengers *inside* the car. In the former case, the kinetic energy of the car is “inside” the system and counts as an output; however, in the latter case, the kinetic energy of the car is “outside” the system and counts as waste (only the kinetic energy of passengers is counted as useful output). The resulting efficiency metric will be different for each boundary definition. While both calculations are “correct”, it is meaningless to compare efficiency metrics that are based on different boundaries definitions. Furthermore, the system boundary used to account for inputs must remain the same as the boundary used to account for outputs. This is easily accomplished in a simple system, but more difficult in a complex system (see below).

Lastly, *unit consistency* is essential for a well-defined efficiency metric (or any calculation, for that matter). If the basic unit of measurement is not well-defined, then quantitative analysis becomes impossible.

Let us now apply these concepts to the neoclassical treatment of decoupling. In abstract form, the idea is seductively simple: the economy grows progressively more efficient at transforming resources into output, and we can measure this efficiency in terms of an output/input ratio. When applied to energy, this becomes the GDP output per unit of energy (or the inverse, the energy intensity of GDP).

However, when we move from theory to the real world, the neoclassical notion of decoupling is beset by fundamental epistemological difficulties. First, we encounter a problem of flow consistency. When following energy inputs, the only way that flow consistency can be maintained is by adopting the Ayres–Warr notion of efficiency, where “output” is measured in energetic terms as useful work (although I would argue that useful work still constitutes an “input”). If we adopt this approach, we find that energy conversion technology has become more efficient, but the growth of this efficiency has strict upper bounds set by the laws of thermodynamics. Thus, the neoclassical notion of perpetual decoupling becomes impossible.

Unlike the Ayres–Warr method, the conventional approach to the measurement of economic efficiency is to abandon the flow of energy after it enters the economy and to instead shift to a notion of “economic output”. By doing so, we lose flow consistency. On the input side we track energy flows, but on the output side we track “economic output”, which consists partly of biophysical flows (the production of goods), but also of human activity (services). It is far from clear what aggregating such disparate phenomena means. By losing flow consistency, we lose the uniqueness of our measure, since there are many possible ways that “economic output” might be defined. For instance, it might make sense to define all services as *inputs*, thereby counting only the production of goods as output. I have argued that from a biophysical standpoint, the only “output” of the entire economy is its waste. All internal outputs are destined to become inputs to other process. Thus, the very notion of an “economic output” is ill-defined.

We also run into problems with boundary definitions. For instance, when measuring energy inputs, the “economic system” is implicitly defined to include any human activity that involves the use of energy (meaning *all* human activity). However, the conventional approach to measuring outputs relies on a different boundary definition. By measuring “economic output”, we are concerned only with the subset of human activity that is *monetized*. Thus, the well-paid banker produces an “output” but the unpaid housewife does not. When we compare “economic output” with energy inputs, we are actually using different boundary definitions of the economic system (monetized human activity vs. all human activity). This inconsistency undermines the validity of our efficiency metric.

Lastly, there is no objective unit on which to measure “economic output”. As discussed extensively in Sect. 2.3, real GDP is plagued by a fundamental instability in its basic unit (price), meaning unit consistency is not maintained.

Thus, when moving from neoclassical theory to the real world, our ability to measure decoupling is undermined by serious (and I would argue, insurmountable) epistemological difficulties. The conventional measure of decoupling—the energy intensity of GDP—fails all three conditions for an effective efficiency metric. Thus, any evidence for decoupling that is provided by this metric should be met with

appropriate scepticism. As such, I argue that the neoclassical notion of decoupling is untestable.

## 2.6 Conclusions: Monetary Value as a Feedback

I have proposed that we abandon the conventional approach of using monetary value to measure output. Instead, I argue that we should treat monetary value as a *feedback device* that controls the flow of biophysical throughput. When we undertake this paradigm shift, the evidence for decoupling disappears. Indeed, we find that kWh-deflated GDP is strongly coupled with changes in the flow of useful work. This does not mean, however, that financial constraints are the ultimate arbiters of biophysical flows. Rather, I argue that financial constraints are a manifestation (in feedback form) of more fundamental problems.

In my opinion, there are three ways in which energy inputs (or any other resource input) to the economy may be constrained:

1. *Biophysical scarcity*: Our ability to consume energy depends fundamentally on both the quantity and quality (i.e., EROI) of an energy resource.
2. *Technological capacity*: Exosomatic energy resources cannot be exploited by humanity without the use of technology that transforms this energy into a form usable by humans (useful work).
3. *Social organization/coordination*: In a complex society, the mobilization of an energy throughput requires the coordination of many individuals. If such coordination cannot be mustered, energy throughput will be constrained.

If anything can be garnered from the study of economic history, it is that when the economy bumps up against one or more of the above three constraints, individuals rarely perceive the truth of what is occurring. Instead, such constraints are universally perceived as a *financial* problem.

We can understand this by way of an analogy with the human body. The body is a complex system with many feedback mechanisms, some of which are conscious. Conscious feedback manifests itself as a “feeling”. When the body requires more energy, we “feel” hungry. When the immune system is under attack, we “feel” unwell. When the body requires time for recuperation, we “feel” tired. In all cases, the “feeling” is a sensory manifestation of a deeper physiological issue.

In the case of the economy, financial constraints are not a “cause” of economic problems; instead they are a manifestation of a deeper biophysical/social issue. The difference, however, is that in the human sensory system, feedback is *qualitative*. The advantage to this qualitative feedback is that we are able to distinguish between different “feelings”, allowing us to respond appropriately. In the case of the economy, monetary feedback is *quantitative*. This has the advantage of being very “precise”, but the disadvantage is that different constraints become indistinguishable when expressed in feedback form.

This leads to a peculiar problem—when faced with constraints on the economy, we often “blame the messenger”. To confirm this, one need only look at the ubiquitous claims made by politicians that scarcity of *money* is the source of the problem. Yet at the level of humanity as a whole, such claims are nonsensical. Money is a creation of the human imagination. To claim that problems arise from a scarcity of money is equivalent to claiming that the stars in the sky cannot be counted because there are not enough numbers. As Soddy (1926) long ago noted, the creation of money is bound not by the laws of thermodynamics but by the laws of mathematics—hence its creation has no upper bound. Therefore, when we witness a financial constraint (and we witness them all the time) we must insist that this is actually a *barometer* for a more fundamental process that is occurring.

We can use the empirical results of this chapter to elucidate this principle. My contention is that kWh-specific purchasing power represents a rough proxy for our ability to finance the consumption of useful work. An increase in this indicator means that useful work becomes “cheaper”. Our response to this signal is to consume more useful work (or to accelerate the growth of this consumption). Alternatively, when kWh-specific purchasing power decreases, useful work becomes “dearer”. Our response to this signal is to consume less useful work (or to slow the growth of our consumption).

Yet the great historical changes in useful work growth rates are not “caused” by financial feedback. Such feedback is ultimately an indicator of either a biophysical, technological, or social constraint (or a combination of the three). Our investigation of the productivity of the mining sector can give us further insight. Mining sector biophysical labor productivity is a joint outcome of biophysical, technological, and social constraints. Better technology will act to increase productivity, while declining resource quantity (and quality) will have the opposite effect. Social constraints also play a role. For instance, the 1970s Organization of the Petroleum Exporting Countries (OPEC) embargo prompted a rapid rise in US drilling rates without a corresponding increase in oil production (Guilford 2011). The result was that productivity declined greatly, but recovered once the embargo ended and drilling intensity relaxed.

Empirical evidence shows that the productivity of the mining sector acts as an arbiter between the price of fossil fuels and the nominal value of the economy. Thus, what is manifested as a financial phenomenon (the “cheapness” or “deariness” of fossil fuels) is actually a reflection of a very concrete biophysical reality—our ability to harvest fossil fuels.

In this chapter, we found evidence of a stable and long-term *coupling* between the consumption of energy and the ability of the economy to finance this consumption. This gives strong support for the hypothesis that monetary value functions as a feedback device. This result has important implications for a biophysical growth theory. Given the strong historical evidence, we can expect that future energy constraints will appear as financial constraints. Indeed, this may already be occurring. The strength of a biophysical growth theory will lie in its ability to demonstrate that the ultimate cause of these financial problems has little to do with money and everything to do with biophysical reality.

## References

- Ayres R, Warr B (2009) *The economic growth engine: how energy and work drive material prosperity*. Edward Elgar Publications, Northampton
- Blas J, Wallis W (2014) Nigeria almost doubles GDP in recalculation. *Financial Times*
- British Petroleum (2012) BP statistical review of world energy. British Petroleum, London
- Bureau of Labor Statistics (2010) Frequently asked questions about hedonic quality adjustment in the CPI. <http://www.bls.gov/cpi/cpihqanda.htm>. Accessed 24 June 2013
- Carter SB, Gartner SS, Haines MR, Olmstead AL, Sutch R, Wright G (2006) *Historical statistics of the United States: Millennial Edition*. Cambridge University Press, Cambridge
- CEC - Commission of the European Communities (2002) Sixth community environment action programme. Off J Eur Union L242:0001–0015
- Chandler D (2010) Explained: The carnot limit. <http://newsoffice.mit.edu/2010/explained-carnot-0519>. Accessed 27 May 2014
- Daly H (1997) Georgescu-Roegen versus Solow/Stiglitz. *Ecol Econ* 22(3):261–266
- Giampietro M, Mayumi K, Sorman A (2012) *The metabolic pattern of societies: where economists fall short*. Routledge, New York
- Giampietro M, Mayumi K, Sorman A (2013) *Energy analysis for a sustainable future: multi-scale integrated analysis of societal and ecosystem metabolism*. Routledge, New York
- Griliches Z (1992) Introduction to “output measurement in the service sectors”. In: *Output measurement in the service sectors*. University of Chicago Press, Chicago, pp 1–22
- Guilford MC, Hall C, O'Connor P, Cleveland C (2011) A new long term assessment of energy return on investment (EROI) for US oil and gas discovery and production. *Sustainability* 3(10):1866–1887
- Jerven M (2012) Lies, damn lies and GDP. *The Guardian*
- Jerven M (2014) *Economic growth and measurement reconsidered in Botswana, Kenya, Tanzania, and Zambia, 1965–1995*. Oxford University Press, Oxford
- King CW, Hall C (2011) Relating financial and energy return on investment. *Sustainability* 3(10):1810–1832
- Meadows DH (1972) *The limits to growth: a report for the Club of Rome's project on the predicament of mankind*. New American Library, New York
- Meadows DH (2012) The energy & finance systems: Mismatched dynamics. In: *Money, debt, and energy: Biophysical realities for fairytale economies*, Burlington, VT. <http://www.uvm.edu/jdericks/BPE4/p2s1.pdf>. Accessed 16 Jan 2013
- Nitzan J (1992) *Inflation as restructuring. a theoretical and empirical account of the US experience*. PhD thesis, McGill University
- OECD (2003) *Environmental indicators: development, measure, and use*. OECD Environment Directorate Environmental Performance and Information Division. Paris
- Sakellaris P, Wilson DJ (2004) Quantifying embodied technological change. *Rev Econ Dyn* 7(1):1–26
- Sherwood MK (1994) Difficulties in the measurement of service outputs. *Mon Labor Rev* 117(3):11–19
- Smith GD, Hart C, Upton M, Hole D, Gillis C, Watt G, Hawthorne V (2000) Height and risk of death among men and women: aetiological implications of associations with cardiorespiratory disease and cancer mortality. *J Epidemiol Community Health* 54(2):97–103
- Soddy F (1926) *Virtual wealth and debt: the solution of the economic paradox*. George Allen & Unwin, London
- Solow RM (1974) The economics of resources or the resources of economics. *Am Econ Rev* 64(2):1–14
- Steindel C (1995) Chain-weighting: the new approach to measuring GDP. *Current Issues in Economics and Finance*, Federal Reserve Bank of New York 1(9)
- Stern D (2004) Economic growth and energy. *Encyclopedia Energy* 2:35–51

- Stiglitz J (1974) Growth with exhaustible natural resources: efficient and optimal growth paths. *Rev Econ Stud* (Symposium on the Economics of Exhaustible Resources) 41:123–137
- UN (1977) Guidelines on principles of a system of price and quantity statistics. Department of Economic and Social Affairs, Statistical Papers M(59)
- Warr B (2009) REXS database. <https://sites.google.com/site/benjaminwarr/the-economic-growth-engine/rexs-database>. Accessed 9 April 2012
- Warr B, Ayres R (2012) Useful work and information as drivers of economic growth. *Ecol Econ* 73:93–102

## Chapter 3

# Distribution

*The produce of the earth ... is divided among three classes of the community ... To determine the laws which regulate this distribution, is the principal problem in Political Economy.*

– David Ricardo (1817)

As David Ricardo's statement makes clear, classical political economists were keenly interested in issues of economic distribution. Within political economy, basic explanations of distribution fell into two broad approaches—one that appealed to conflict and one that appealed to the harmony of the market. The conflict school is best characterized by the theories of Karl Marx (1867), who theorized that distribution was the result of worker vs. capitalist class struggle. The harmony school can be traced back to Adam Smith (1776). While his work contained elements of both harmony and conflict, it was Smith who put forward the basic tenet of the invisible hand: the self-serving actions of individuals, operating in a free market, function to benefit society as a whole.

The so-called *marginal revolution* of the late nineteenth century, spearheaded first by Menger (1871), Jevons (1879), and Walras (1896) and later by Marshall (1890), Wicksteed (1894), and Clark (1899), gave mathematical rigour to Smith's qualitative theories. This new-classical, or *neoclassical*, approach theorized that under conditions of perfect competition and constant returns to scale, the size of one's income was proportional to the marginal productivity of one's labor and/or property. Milton Friedman (1962) summarizes this dictum succinctly: "to each according to what he and the instruments he owns produces."

Because income is theorized to be proportional to productivity, when adopting the neoclassical approach, one is much more likely to conclude that existing income distribution is "fair."<sup>1</sup> The result of this theoretical worldview is that distribution effectively becomes a nonissue. Thus, as neoclassical theory began to dominate economic thinking and teaching, interest in distributional issues waned. Atkinson and Bourguignon note that:

---

<sup>1</sup> The catch is that one's productivity depends upon one's initial property endowment, which could be decidedly unequal. However, questions surrounding the equity of this initial endowment are generally relegated to disciplines outside economics (i.e., sociology).



[I]n the second half of the century, there were indeed times when interest in the distribution of income was at a low ebb, [with] economists appearing to believe that differences in distributive outcomes were of second order importance compared with changes in overall economic performance. (2001, p. 7265)

Nowhere was this disinterest more clear than in the development of neoclassical growth theory, which effectively swept away issues of distribution entirely. Bertola et al. write:

While early growth models in the post-Keynesian tradition were still strongly concerned with distributional issues, subsequent “new-classical” theoretical developments removed distribution from the set of macroeconomic issues of interest ... The distribution of income and wealth across consumers was viewed as a passive outcome of aggregate dynamics and market interactions, and little attention was paid to feedback effects from distribution into growth and other macroeconomic phenomena. (2006, p. x) [emphasis added]

Thus, permeating neoclassical growth theory is the assumption that distribution is unimportant to growth. But is such an assumption justified? In order to answer this question, this chapter investigates the link between growth and three types of distribution:

**Functional:** Distribution by income type (i.e., rent, profit, wages, etc.)

**Debtor/Creditor:** Distributional claims on future production

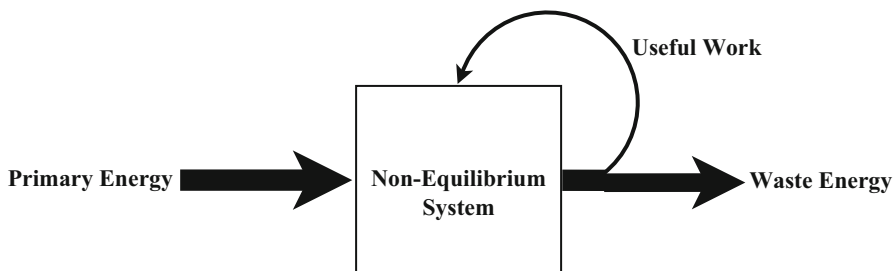
**Personal:** Distribution by income size (i.e., income inequality)

In this chapter, each of these assumptions is subjected to empirical test. In all cases, the results show that growth and distribution are fundamentally linked, meaning neoclassical growth theory assumes away a core component of the growth process.

### 3.1 Measuring Growth

Before we can investigate the link between distribution and growth, we must first settle on a method for measuring the scale of the economy. This matter is not trivial. The notion of “scale” has *no meaning* without an observer first stipulating both the criteria and methodology by which “scale” is to be defined and measured. Different choices will yield radically different results, all of which are “rigorous” and “correct”. Giampietro et al. call this the “epistemological predicament associated with purposive quantitative analysis . . . the observer always affects what is observed when defining the descriptive domain” (2006, p. 307).

Economists make the preanalytic decision to measure economic scale in terms of monetary value. However, this approach leads to a quagmire of difficulties stemming from the fact that the basic unit of measurement (price) undergoes nonhomogeneous changes through time. Furthermore, the quality of output changes over time, and there are no objective means for measuring such qualitative change. Due to these problems, I abandon the real gross domestic product (GDP) approach for the remainder of this book. Instead, I propose focusing on a *biophysical* definition of scale. While there



**Fig. 3.1** A conceptual model of a nonequilibrium system

are many possible ways that biophysical scale might be defined, I choose to focus on energy flows.

Astrophysicist Eric Chaisson writes that “energy, the ability to do work, is the most universal currency known in natural sciences” (2005, p. 21). Energy flows are a prerequisite for the existence of any complex system. Without such flows, the second law of thermodynamics stipulates that all roads lead to equilibrium. Individual organisms, ecosystems, and human societies can all be treated as different types of nonequilibrium systems that maintain their complexity through energy dissipation (Kondepudi and Prigogine 1998).

The basic conceptual model of a nonequilibrium system is displayed in Fig. 3.1. The system exists because of some sort of energy flow. Most of this energy simply passes through the system, but some of it (useful work) is used to maintain/increase the internal order of the system. In the case of humans, the energy flow is the chemical energy in food, the waste energy is radiative heat plus the chemical energy in undigested excrement, and the useful work is the cellular metabolism that maintains our structure and allows us to grow/reproduce.<sup>2</sup>

As an analogy to biological metabolism, we can refer to human society (or the economy) as having a *social metabolism* (Giampietro et al. 2012; Martinez-Alier 2007). In specific form, an individual’s metabolism is wildly different than an entire society’s social metabolism. However, on a conceptual level they are identical: some form of useful work must constantly be done in order to counterbalance the forces of entropy. If either system is to grow, more useful work must be done.

For the remainder of the book, I use Ayres and Warr’s estimates of useful work as a measure of the biophysical scale of the economy. This data is freely available from Warr’s REXS database (Warr 2009). When useful work data is unavailable (for instance, on an international or subnational scale) I simply use primary energy consumption. I make no arguments that this approach is better or more accurate than using real GDP as a measure of scale—only that it is more physically meaningful and requires fewer subjective decisions. I believe that a biophysical growth theory

<sup>2</sup> We ignore the added complication that energy from food is also used to do work on systems that are external to the body (i.e., manual labor).

should give insight into how changes in energy consumption affect society. By using energy as our measure of scale, such insight is automatically embedded in the theory.

## 3.2 Functional Distribution

To discuss the role of functional distribution in neoclassical growth theory, we return to the Cobb–Douglas production function used in the Solow–Swan model:

$$Y = AL^\beta K^\alpha \quad (3.1)$$

Recall that  $\alpha$  represents the income share of capital and  $\beta$  the income share of labor, and that both are theorized to be proportional to each factor’s marginal productivity. While  $\alpha$  and  $\beta$  could conceivably be treated as *variables*, in practice they are treated as *constants*. This tradition has its origins in the work of Cobb and Douglas (1928), who showed that fixed exponents could be used to model historical production. The constancy of factor shares was later formalized by of Nicholas Kaldor (1957) who put forward a list of six stylized facts about economic growth, one of which was the historical tendency for capital and labor income shares to remain approximately constant over time (about 1/3 and 2/3 respectively). In neoclassical growth theory, it has become standard practice to assume that functional income distribution is constant. From this, the following assumption is implicit:

**Functional Distribution Assumption:**

Changes in functional distribution are unrelated to growth.

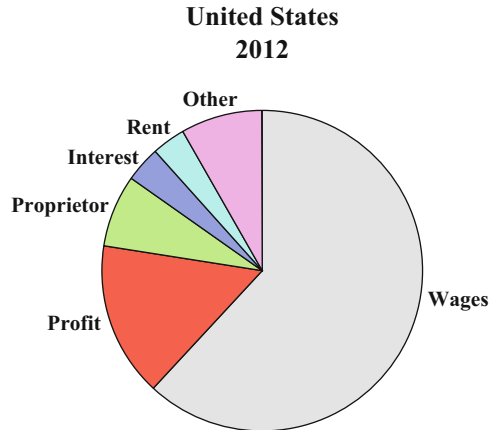
As a rough first estimate, functional distribution is *approximately* constant over time; however there are major oscillations around the long-term average, especially if we disaggregate capitalist income into its constituent parts. For instance, over the period 1929–2012, profit represented an average of 10.2 % of the US national income (data from BEA Table 1.12). However, the relative standard deviation in this share was a nontrivial 27 %, meaning variability was roughly 1/3 as large as the mean. If we treat profit’s share as a constant, then we implicitly assume that these large deviations from the long-term average are unimportant.

Rather than using the capital-labor binary of neoclassical growth theory, I used the five types of income represented in the national accounts: *wages*, *profit*, *proprietor*, *interest*, and *rent*. Wages represent the income of all employees, both salaried and waged. Profit represents the income of owners of incorporated businesses, while proprietor represents the income of unincorporated businesses (mostly self-employment and partnerships). Interest represents the income of creditors (owners of debt). Lastly, rent represents the income of owners of specific property (natural resources, land, intellectual property, etc.).<sup>3</sup> The US 2012 income “pie” is displayed in Fig. 3.2.

---

<sup>3</sup> The rent category is complicated by the accepted practice of treating home ownership as a *business* activity. Thus, the BEA calculates an *imputed* rent for all owner-occupied buildings. To whom this rent is actually paid remains unclear.

**Fig. 3.2** The income pie.  
(Sources BEA Table 1.12)



Our goal is to investigate how distribution between these five income-types is related to the growth of useful work. My methodology is to compare relative *changes* in functional distribution with changes in useful work production. This approach is an adaptation of the methodology used by Bichler and Nitzan (2013) for a similar investigation into the link between distribution and unemployment. Thus, we compare the annual rate of change in the share of each income type to the annual rate of change of useful work. Equation 3.2 shows a sample calculation for profit ( $\pi$ ). Here  $Y$  refers to national income,  $U_{pc}$  refers to useful work per capita, and  $\hat{\phantom{x}}$  denotes growth rate. Correlations for this procedure, carried out across the five income types, are shown in Table 3.1.

$$\left[ \frac{\pi}{Y} \right] \hat{\phantom{x}} \iff \hat{U}_{pc} \quad (3.2)$$

The results show that there is significant correlation between the growth of useful work and changes in the income share of both profit and interest. All other income types show much smaller correlation. Profit is the only income type to show significant *positive* correlation with growth. Thus, profit experiences an *increase* in share during periods of expansion (Fig. 3.3). Interest, on the other hand, experiences relative growth during periods of *contraction*.

What we are seeing in this data is mostly fluctuations due to business cycles; however, the results are not as straightforward as they might seem. Because of the importance of profit to capitalist economies, one is tempted to assume that the *share* of profit is at a maximum during the peak of the business cycle (here defined as a peak in useful work growth rates). This would mean that maximum profit equates with maximum growth. However, this turns out to be untrue. The correlation between profit's share of national income and useful work growth rates is virtually nil ( $R^2 = 0.002$ ). It is not the *share* of profit, but its *rate of change* that is correlated with growth. Useful work growth is at its peak when the *rate of change* of the share of profit is at a maximum.

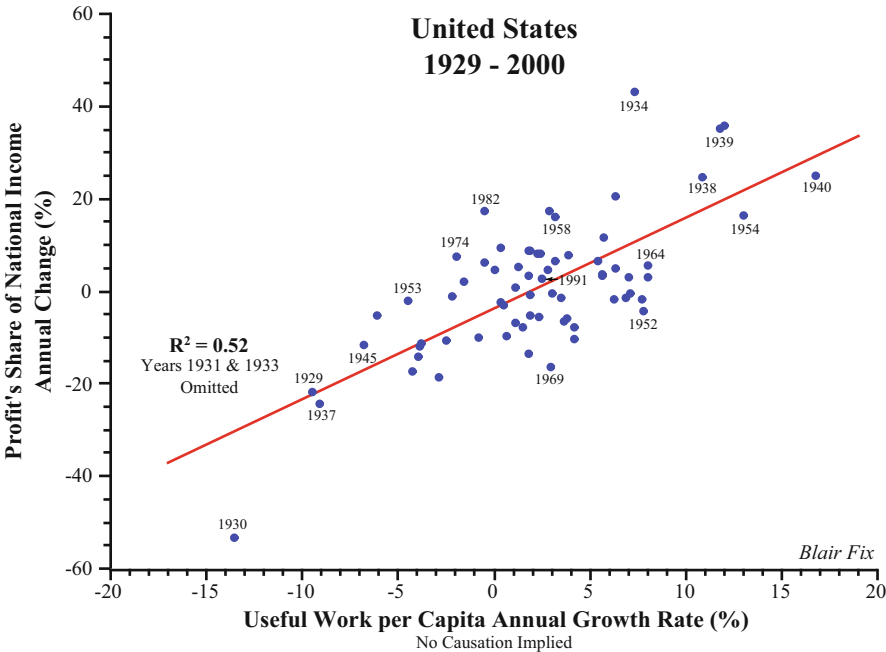
**Table 3.1** Correlation between income redistribution and useful work growth

Income type	$R^2$
Profit	0.52
Interest	(−) 0.28
Rent	(−) 0.16
Wages	(−) 0.15
Proprietor	0.08

Note: Data is for correlation between annual growth rate of useful work per capita and annual change in each factor’s share of national income. Profit correlation excludes the years 1932–1933; when included  $R^2$  drops to 0.23. All income data from Bureau of Economic Analysis, Table 1.12. Data series: Corporate profits (before tax) with IVA and CCAdj, Net interest and miscellaneous payments, Rental income of persons with CCAdj, Compensation of employees, Proprietors’ income with IVA and CCAdj. Data for useful work and US population is from Benjamin Warr’s REXS database

While these results are in many ways perplexing, they do make one thing clear: it is misguided to assume (as neoclassical theory does) that changes in functional distribution are not important to growth. But to be fair to neoclassical growth theory, it is important to recognize that the Solow–Swan model was never intended to give insight into business cycles. However, the focus on stable, “balanced” growth and the avoidance of cyclical fluctuations is problematic in its own right. The current understanding of nonequilibrium systems is that short-term fluctuations play an instrumental role in the creation of order—what physicist Ilya Prigogine calls “order through fluctuations” (Kondepudi and Prigogine 1998).

As the results indicate, fluctuations in functional distribution show significant connection with growth. By ignoring such fluctuations, neoclassical growth theory maintains simplicity by assuming away a key piece of the puzzle. Can a biophysical growth theory explain why fluctuations in profit are so important to growth? The empirical results of Chap. 4 suggest one plausible explanation. It seems that large corporations play a fundamental role in the growth of energy consumption. Given that the primary goal of such corporations is to make a profit, building on our results from Chap. 2, we might hypothesize that profit functions as a feedback mechanism that influences the behavior of business. If business is ultimately in control of the economy, fluctuations in profit would tend to dominate other forms of feedback (wages, proprietor income, etc.). While plausible, an investigation of this hypothesis must be left for future research.



**Fig. 3.3** Annual changes in profit's share of national income vs. annual change in useful work per capita. Note: No causation is implied by the choice of x- and y-axes variables. (Sources: Data for useful work and population from Benjamin Warr's REXS database. Profit and national income data from BEA Table 1.12)

3.3 Debtor/Creditor Distribution

In the preceding section, we saw that changes in the interest share of national income were negatively correlated with the growth of useful work. This raises an important question: since interest accrues to the owners of debt (i.e., creditors), how is debt related to growth?

As Foster and Magdoff (2009) emphasize, debt is a distributional issue. However, it is one that is slightly different from our previous discussion. While interest payments are the present income stream derived from debt, the debt itself represents the total amount owed to the creditor. Debt is thus a distributional claim, made by creditors, on the *future* income of debtors. In this section, we investigate how debtor/creditor distribution is related to biophysical growth.

First, we note that debt is completely absent from the Solow–Swan model, leading to the following implicit assumption:

**Debt Assumption:** Debt levels are unrelated to growth.

In order to understand this assumption, we must trace a long lineage of theory. We begin by looking at the Solow–Swan capital accumulation formula:

$$\dot{K} = sY - \sigma K \quad (3.3)$$

Aside from technological progress, which is unexplained, capital accumulation is the main engine of growth in the Solow–Swan model. In Eq. 3.3,  $\dot{K}$  represents the rate of change of capital (i.e., the rate of capital accumulation),  $s$  represents the savings rate,  $Y$  the quantity of total production,  $\sigma$  the rate of capital depreciation (how quickly capital goods decay), and  $K$  the size of the capital stock. The logic behind this equation is that savings are a form of abstinence from present consumption. If the savings rate is zero, all that is produced is consumed, and no capital is accumulated ( $\dot{K} \leq 0$ ). However, if some actors forgo consumption, this abstinence will take the form of savings ( $s > 0$ ). The portion of production that is saved ( $sY$ ) can then be *invested* in the production of capital goods. So long as total investment ( $sY$ ) is greater than total depreciation ( $\sigma K$ ), capital will *accumulate* ( $\dot{K} > 0$ ).

Thus, it is the savings rate, by and large, that determines the level of capital accumulation in the Solow–Swan model. Paradoxically, this emphasis on savings exists without any reference to the role of *money*. This absence is indicative of the neoclassical view that money is a neutral facilitator of exchange. Parguez and Seccareccia elaborate:

The starting proposition of the neoclassical approach ... is that money is to be *introduced* within an otherwise pre-existing pure barter economy ... All monetary transactions are *de facto* of the same nature as barter exchange with a *numéraire* money having been introduced to it. (2000, p. 113)

In the Solow–Swan model, the savings rate represents *physical* abstinence from consumption, but is accounted for in terms of the *monetary value* of savings. Neoclassical theory assumes that these two forms are equivalent (Nitzan and Bichler 2009). Since money plays a passive bookkeeping role, there is no need for it in neoclassical growth theory. Milton Friedman summarizes the neoclassical position: “nothing is so unimportant as the quantity of money expressed in terms of the nominal monetary unit” (1969, p. 1).

However, we are left with a sticky question: if the value of output grows, must not the amount of money in existence grow as well? How is this accomplished? In neoclassical theory, the money supply is an exogenous parameter determined by the central bank. This institution first creates the monetary base, and then sets the fractional reserve rate, which determines the rate at which commercial banks can “multiply” this base. Thus, in neoclassical theory, money creation is an exogenous and “neutral”<sup>4</sup> parameter that is controlled by a “benevolent monetary dictator” (Arestis and Sawyer 2001, p. 201), which in the US, is the Federal Reserve Bank.

---

<sup>4</sup> On this topic, Keynes joked, “The [theoretical] conditions required for the “neutrality” of money ... are, I suspect, precisely the same as those which will ensure that crises do not occur” (1971, p. 410).

This is in marked contrast to the “endogenous” money theory of the post-Keynesian tradition, in which debt/credit relations are theorized to play a central role in the creation of money. In endogenous money theory, money is created whenever credit is advanced. For instance, a farmer may obtain grain seed on credit, whereby the seller holds an IOU from the farmer. This extension of credit creates money (the IOU). However, its viability (beyond the two-person relationship in which it was created) depends on the degree to which other people trust the farmer’s ability to repay his debts. If the farmer commands great esteem, the IOU can circulate widely.

In principle, anyone can create money in this manner. In practice, however, commercial banks have almost completely monopolized this process. According to endogenous money theory, when advancing credit, banks do not lend savers’ money; rather, they create money out of nothing and register it as an outstanding debt. Interestingly, the Bank of England itself recently endorsed this view:

In the modern economy, most money takes the form of bank deposits. But how those bank deposits are created is often misunderstood: the principal way is through commercial banks making loans. Whenever a bank makes a loan, it simultaneously creates a matching deposit in the borrower’s bank account, thereby creating new money. (McLeay et al. 2014, p. 1)

Endogenous money theory has important implications regarding debt. Most importantly, the growth of the money supply is dependent on increasing the total outstanding debt. The money supply, in turn, is related to the value of total production ( $Y_n$ ) by the following identity:

$$MV \equiv Y_n \quad (3.4)$$

This restatement of Irving Fisher’s (1911) famous equation of exchange<sup>5</sup> tells us that if the value of production increases, either the money supply  $M$  or the transaction “velocity” of money in final expenditures ( $V$ ) must increase. Since the latter is usually assumed to be constant, it follows that increases in the value of production must be accompanied by an increase in the amount of money. According to endogenous money theory, this requires an increase in debt, leading to the conclusion that the growth of debt is inherently important to the growth of value.

Within the neoclassical tradition, on the other hand, debt is regarded as passive and has no involvement in the creation of money (see Hudson 2011 for a good review of the neoclassical treatment of debt). Rather, debt is explained in terms of the *loanable funds* model, in which savers loan their funds to borrowers (and banks act as intermediaries)<sup>6</sup>. In his textbook on macroeconomics, Gregory Mankiw explains:

The supply of loanable funds comes from people who have some extra income they want to save and lend out. This lending can occur directly, such as when a household buys a bond from a firm, or it can occur indirectly, such as when a household makes a deposit in a bank,

---

<sup>5</sup> Fisher’s original equation is  $MV \equiv PQ$ , where  $Q$  is the quantity of transactions and  $P$  is the average price level. The right-hand side is equivalent to nominal GDP.

<sup>6</sup> One might question how valid it is to treat banks as mere intermediaries when their business model becomes heavily geared towards speculative securities.



which in turn uses the funds to make loans. In both cases, saving is the source of the supply of loanable funds. (2011, p. 268)

In the loanable funds model, changes in debt levels (i.e., debt inflation or deflation) are theorized to be unimportant. Ben Bernanke summarizes this position:

Debt-deflation represent[s] no more than a redistribution from one group (debtors) to another (creditors). Absent implausibly large differences in marginal spending propensities among the groups . . . pure redistributions should have *no significant macro-economic effects*. (1994, p. 17) [emphasis added]

Interestingly, the Great Recession has precipitated a rather fierce debate between Steve Keen and Paul Krugman on the merits of endogenous and exogenous money theory (Keen argues for the former and Krugman the latter). Amidst the debate, Krugman dismisses the importance of debt:

Ignoring the foreign component, or looking at the world as a whole, we see that the overall level of debt makes no difference to aggregate net worth—one person’s liability is another person’s asset. It follows that the level of debt matters only if the distribution of net worth matters, if highly indebted players face different constraints from players with low debt. (Krugman 2012, p. 146)(cited in Keen 2014, p. 5)

We thus have two diametrically opposed schools of thought: the neoclassical school suggests that growth and debt levels are unrelated, while the endogenous money school suggests that the two ought to be fundamentally related. As is common in economics, the debate between these two schools has largely played out on a theoretical level. In this book, however, we are concerned with empirical evidence.

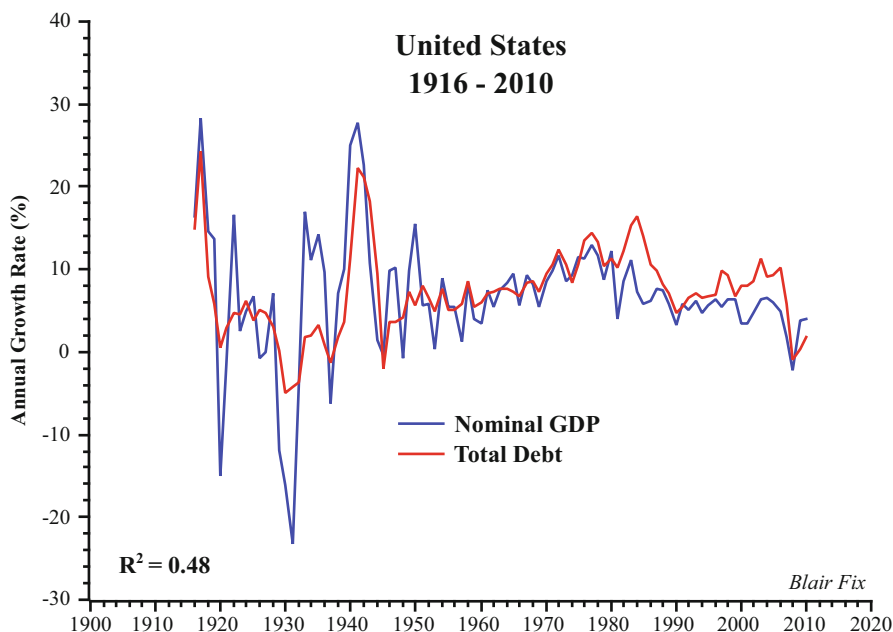
We begin by looking at the relation between the growth of nominal GDP and the growth of total debt (Fig. 3.4). The evidence is clear: the growth rate of the value of production is highly coupled with the growth rate of total debt. This lends good empirical support for endogenous money theory. However, we must be careful not to conflate the growth of value (nominal GDP) with the growth of biophysical scale (useful work). The two are entirely different. Noting this, we still have no indication of the relation between debt and biophysical growth.

We can investigate this relation by dividing total debt ( $D$ ) by nominal GDP ( $Y_n$ ), giving us the “debt-to-GDP” ratio. We then compare the annual percentage change in this ratio to the annual growth rate of useful work per capita:

$$\left[ \frac{D}{Y_n} \right] \Longleftrightarrow \hat{U}_{pc} \quad (3.5)$$

The results, shown in Fig. 3.5, are unequivocal. Changes in outstanding debt, measured against the value of current production, exhibit a tight, *negative* correlation with the growth of useful work. We can immediately conclude that the Solow–Swan assumption that debt levels are unimportant to growth is contradicted by the empirical evidence.

Since debt carries units of \$ and nominal GDP carries units of \$/year, the debt-to-GDP ratio carries units of years. We can interpret this as the hypothetical time required to pay existing debts if the total value of present-day production were devoted to this

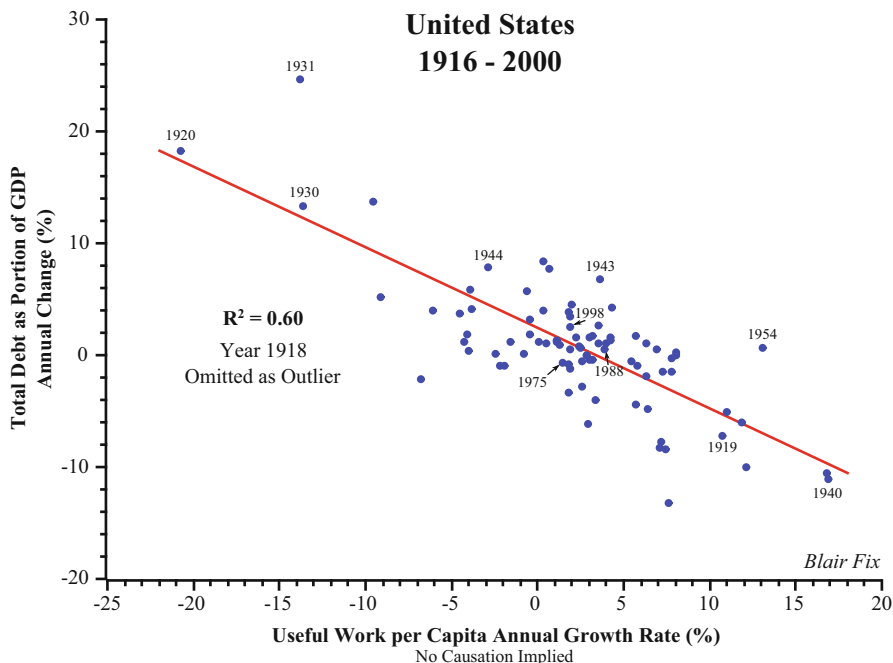


**Fig. 3.4** Annual changes in total debt and nominal gross domestic product (GDP) (Sources: see Fig. 3.5)

task. The shorter this time period, the more feasible is debt repayment. Figure 3.5 indicates that the ability to repay debt is highly coupled to biophysical growth. When the biophysical scale of the economy expands, debt repayment becomes more feasible. Conversely, when the economy undergoes biophysical contraction, debt repayment becomes less feasible. Note, however, that we cannot determine causation from this data: debt levels could be driving growth rates, or vice versa. Or, the two might exist in a coupled feedback relation in which causation is circular. Without a deeper investigation (which exceeds the scope of this book) causation must remain ambiguous.

Note that the relation between debt and useful work is not simplistic. If we break the correlation between useful work and debt-to-GDP into pre- and post-1970 eras, we get very different results. For the period 1916–1970, the correlation is strong ( $R^2 = 0.66$ ); however, for the period 1971–2000, correlation between useful work and debt-to-GDP is nonexistent ( $R^2 = 0.005$ ). Thus, there is a clear structural change that took place around 1970. This corresponds with the end of the gold standard and the introduction of neoliberal policies geared towards trade liberalization. It also coincides with a peak in the labor productivity of the mining sector (Fig. 2.10), a peak in US oil production, and the explosion of the US trade deficit. While interesting, I leave an investigation into this structural change for future work.

The relation between debtor/creditor levels and biophysical growth is evidently complex. While we can safely conclude that neoclassical growth theory's dismissal



**Fig. 3.5** Annual change in debt-to-GDP ratio vs. annual change in useful work per capita. (Note: No causation is implied by the choice of x- and y-axis variables. Sources: Data for useful work and population from Benjamin Warr's REXS database. Sources: Total debt (public and private) for 1916–1945 from HSUS Table Cj870. Debt data for 1945–2000 from US Federal Reserve, credit market instruments liabilities: sum of domestic nonfinancial sectors (Z1/Z1/LA384104005.A) and financial business (Z1/Z1/LA794104005.A). Nominal GDP from HSUS Table Ca10 (1916–1928) and BEA Table 1.1.5 (1929–2000))

of debt is not justified by the evidence, we cannot substitute this dismissal with a simple, deterministic relation.

### 3.4 Personal Distribution

In his textbook on economic growth theory, Daron Acemoglu notes that in order to justify the use of an aggregate production function, neoclassical growth theory must assume that the economy permits the use of a *representative household*. As Acemoglu puts it, this means that the “preference (demand) side of the economy can be represented as if there were a *single household* making the aggregate consumption and saving decisions” (2008, p. 149, emphasis added). This amounts to the assumption that the group of all households can be meaningfully represented by its average.

However, when we represent a group by its average, we lose information. Consider a 100 household economy with a representative (or average) household income of \$ 50,000. By looking only at the representative household, we cannot distinguish between: (1) an economy in which all 100 households earn exactly \$ 50,000 each; and (2) an economy in which 99 households earn nothing and one household earns \$ 5 million. Thus, the use of an aggregate production function—and accompanying representative household—requires that we ignore personal income distribution and its associated inequalities. We can conclude that the Solow–Swan model, which has an aggregate production function at its heart, requires the following implicit assumption:

**Personal Distribution Assumption:**

Income inequality is unrelated to growth.

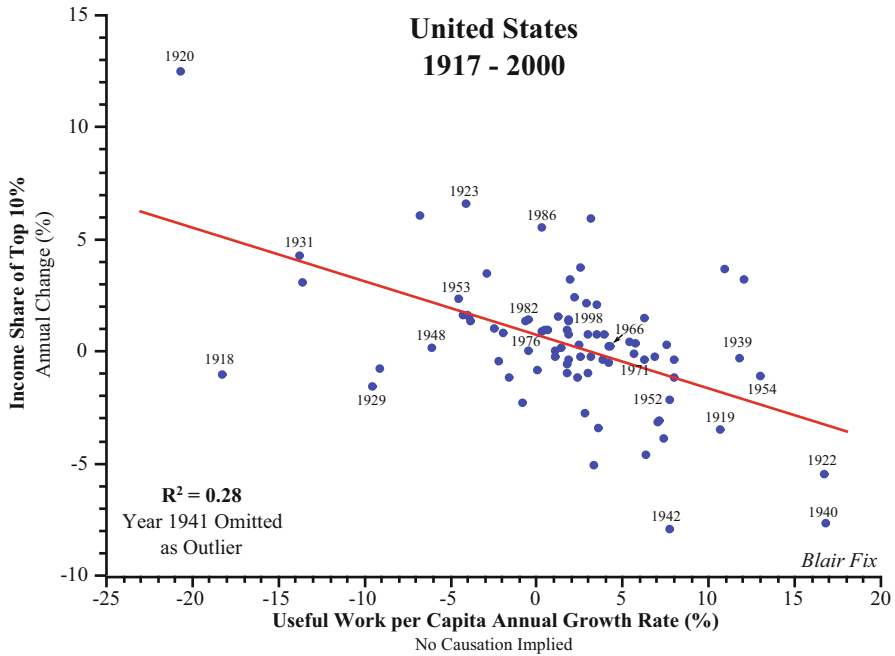
While the Solow–Swan model assumes away any link between personal income distribution and growth, the equity-growth connection (or lack thereof) remains highly contested in economic theory. Two opposing schools exist that advance mutually contradictory theories. The “trickle-down” (supply-side) school argues that income inequality is good for growth, while the “underconsumption” (demand-side) school argues that inequality is harmful to growth. Both theories hinge on the assumption that the “marginal propensity to save” (the change in savings per change in income) differs between income classes.

The notion of differential savings rates among income classes can be traced back to post-Keynesian economists Nicholas Kaldor and Luigi Pasinetti. Kaldor (1955) argued that the marginal propensity to save should be greater among capitalists than among workers, while Pasinetti (1962) argued that the marginal propensity to save was an increasing function of income size (regardless of functional income type). Since the rich have a greater marginal propensity to save than the poor, the trickle-down school argues that an increase in inequality will result in increased savings. According to neoclassical growth theory, the savings rate determines the rate of capital accumulation, which is, in turn, one of the main engines of growth. Thus, it follows that greater income inequality should ultimately yield an increase in growth.<sup>7</sup>

Underconsumption theory turns this logic on its head by focusing on the role of consumption (rather than savings) in driving growth. This school asserts that consumption plays a crucial role in driving demand for the goods and services that are produced. It is demand (not savings) that drives growth. Since lower income classes have a greater “marginal propensity to consume” (the reverse of the propensity to save), a decrease in income inequality should lead to an increase in effective demand and, hence, greater growth. This approach is most often associated with Keynesians such as Kalecki (1971) and Keynes himself, and is now argued for prominently by Robert Reich (2013) and Joseph Stiglitz (2013).

---

<sup>7</sup> As Keynes pointed out, this argument falls apart if savings are hoarded rather than invested.



**Fig. 3.6** Annual change in top 10 % income share vs. annual change in useful work per capita. (Note: No causation is implied by the choice of x and y-axes variables. Sources: Data for useful work and population from Benjamin Warr's REXS database. Income share of top 10 % from The World Top Incomes Database (Alvaredo et al. 2013))

Combining underconsumption and trickle-down theory with the Solow–Swan model assumptions, we cover the full range of possible connections between inequality and growth: harmful, beneficial, and unrelated. In order to test this link, we apply a methodology similar to the previous sections. We compare the percentage rate of change of inequality—quantified in terms of the fractional income share of the top 10 % ( $f_{T10\%}$ )—with the percentage rate of change of useful work per capita:<sup>8</sup>

$$\hat{f}_{T10\%} \Longleftrightarrow \hat{U}_{pc} \quad (3.6)$$

The results, shown in Fig. 3.6, demonstrate that personal income *redistribution* (i.e., the rate of change of income distribution) is correlated with biophysical growth. Although the correlation is lower than for functional and debtor/creditor distribution, we can still conclude, contrary to the assumptions made by the Solow–Swan model, that personal income distribution is related to growth. Furthermore, we can state

<sup>8</sup> Note that we are comparing the *change* in income inequality with the *change* in useful work per capita (i.e., first derivatives). If we compare the *level* of income inequality (not its rate of change) with the rate of change of useful work per capita, there is no correlation at all ( $R^2 = 3.0 \times 10^{-4}$ ).

that the evidence does not support trickle-down theory: rates of change of inequality are *negatively* correlated with growth. Thus, there is no evidence that allowing the wealthy to increase their share of income increases biophysical growth.

Having eliminated two of the three possible relations between inequality and growth, it seems that the evidence must support underconsumption theory. And yet, the matter is more complicated than it appears in Fig. 3.6. If we divide the correlation into two periods, as we did with debt, we find dramatically different results. For the period 1917–1985, the correlation is negative, with an  $R^2$  of 0.28. However, for the subsequent period (1986–2000), the relation turns slightly positive, but with a negligible  $R^2$  of 0.08.

Furthermore, we cannot tell which way the line of causation runs. Underconsumption theory firmly states that causation should run from distribution to growth—a decrease in inequality should stimulate growth. However, we cannot eliminate the alternative that an increase in growth simply makes downward redistribution more *palatable*—that higher taxes on the wealthy are more easily swallowed when useful work grows more quickly. Alternatively, it could be that growth leads to a shortage of labor, thereby increasing the bargaining power of the bottom 90 %. Furthermore, causation could be circular (i.e., feedback related), with both growth and personal income distribution mutually influencing one another. Without further investigation, which exceeds the scope of this book, we must be content with correlation and leave the discussion on causation for a later date.

Despite the complexity of the issue, we can conclude that neoclassical growth theory is not justified in assuming that personal income distribution is unrelated to growth.

### 3.5 Conclusions: Distribution Matters

Through a series of preanalytic decisions, neoclassical growth theory removes distribution from its scope. If this decision is justified, we would expect to find very little correlation between rates of growth and changes in distribution. Yet the empirical evidence presented in this chapter leads to the exact opposite conclusion: biophysical growth (as measured by useful work) is *systematically* correlated with changes in distribution. Clearly the implicit assumption that distribution is unrelated to growth, rests on dubious grounds.

However, if we place neoclassical growth theory in historical context, the decision to ignore distribution makes sense. The Solow–Swan model was formulated in 1956, at the dawn of an era of unprecedented expansion. After World War II, growth became the mantra of the day. Historian J. R. McNeill summarizes:

Capitalists, nationalists—indeed almost everyone, communists included—worshipped at this same altar because economic growth disguised a multitude of sins. Indonesians and Japanese tolerated endless corruption as long as economic growth lasted. Russians and eastern Europeans put up with clumsy surveillance states. Americans and Brazilians accepted vast social inequalities. Social, moral, and ecological ills were sustained in the interest of

economic growth; indeed, adherents to the faith proposed that only more growth could resolve such ills. Economic growth became the indispensable ideology of the state nearly everywhere. (2000, p. 334)[as cited in Daly 2011]

Reviewing the empirical evidence presented in this chapter, it is easy to see why policymakers and economists focused on growth and not on distribution. It is because growth eased distributional issues. Summarizing our findings, we can state that increases in the useful work growth rate are related to:

1. Increases in profit's share of national income.
2. The increased ability of debtors (both public and private) to repay creditors.
3. Decreases in income inequality.

At the turn of the twenty-first century, distributional issues are once again of central interest to economists and policymakers alike. For instance, Thomas Piketty's (2014) book on income inequality has become an unlikely New York Times best seller. The evidence presented in this chapter suggests that distribution and growth are inherently connected—distributional issues are magnified when biophysical growth stagnates. Biophysical scholars have long predicted that peak oil would spell the end of growth (Hubbert 1993; Heinberg 2011; Murphy and Hall 2011). With global oil production currently in a plateau, we should not be surprised that distributional issues are once again at the forefront of public interest.

As stated at the outset of the book, complexity poses a major problem for science. Without a certain amount of idealization, science is impossible. Simplify too far, however, the scientific theory will be of little use. Removing the real-world complexities of economic distribution greatly simplifies the process of explaining growth. Yet, the evidence suggests that this simplification is misguided: the complex business of distribution seems to be inherently connected to biophysical growth. As such, the neoclassical assumption that distribution is unrelated to growth is *insidiously* untrue.

From a Darwinian perspective, the control of resources is one of the major drivers of reproductive success. Although we should not reduce this to social Darwinism, it is naive to think that humans are somehow separate from an otherwise universal resource struggle. From the evidence presented here, it is clear that the growth of energy consumption dampens the distributive struggle. After all, the easiest way to reduce the competition for resources is to increase the resource “pie”. This effectively makes the game “positive sum”: there are relative winners and losers, but overall, *everyone wins*. Yet the peak and decline of energy production threatens to make the game “negative sum”: there will be relative winners and losers, but overall, *everyone loses*. Thus, while predictions are difficult to make, a future decline in energy consumption will almost certainly exacerbate existing distributional problems.

## References

- Acemoglu D (2008) Introduction to modern economic growth. Princeton University Press, Princeton
- Alvaredo F, Atkinson T, Piketty T, Saez E (2013) The world top incomes database. <http://topincomes.parisschoolofeconomics.eu/#>. Accessed 22 Dec 2012

- Arestis P, Sawyer MC (2001) Money, finance and capitalist development. Edward Elgar Publishing, Northampton
- Atkinson A, Bourguignon F (2001) Income distribution. In: International encyclopedia of the social and behavioral sciences (Economics/Public and Welfare Economics). Elsevier, Amsterdam, pp 7265–71
- Bernanke BS (1994) The macroeconomics of the Great Depression: a comparative approach. w4814, National Bureau of Economic Research
- Bertola G, Foellmi R, Zweimüller J (2006) Income distribution in macroeconomic models. Princeton University Press, Princeton
- Bichler S, Nitzan J (2013) Can capitalists afford recovery? Economic policy when capital is power. Working Papers on Capital as Power
- Carter SB, Gartner SS, Haines MR, Olmstead AL, Sutch R, Wright G (2006) Historical statistics of the United States: millennial edition. Cambridge University Press, Cambridge
- Chaisson E (2005) Non-equilibrium thermodynamics in an energy-rich world. In: Kleidon A, Lorenz R (eds) Non-equilibrium thermodynamics and the production of entropy: life, earth, and beyond. Springer, New York, pp 21–31
- Clark JB (1899) The distribution of wealth. Macmillan, New York
- Cobb CW, Douglas PH (1928) A theory of production. *Am Econ Rev* 18(1):139–165
- Daly H (2011) Ecological economics : principles and applications. Island Press, Washington
- Fisher I (1911) The purchasing power of money: its determination and relation to credit interest and crises. Macmillan & Company, New York
- Foster JB, Magdoff F (2009) The great financial crisis: causes and consequences. Monthly Review Press, New York
- Friedman M (1962) Capitalism and freedom. University of Chicago press, Chicago
- Friedman M (1969) The optimum quantity of money. Transaction Publishers, New Brunswick
- Giampietro M, Allen TFH, Mayumi K (2006) The epistemological predicament associated with purposive quantitative analysis. *Ecol Complex* 3(4):307–327
- Giampietro M, Mayumi K, Sorman A (2012) The metabolic pattern of societies: where economists fall short. Routledge, New York
- Heinberg R (2011) The end of growth: adapting to our new economic reality. New Society Publishers, Gabriola
- Hubbert MK (1993) Exponential growth as a transient phenomenon in human history. In: Daly H, Townsend K (eds) Valuing the Earth: economics, ecology ethics. MIT Press, Cambridge, pp 113–126
- Hudson M (2011) How economic theory came to ignore the role of debt. *Real-World Econ Rev* 57:2–24
- Jevons WS (1879) The theory of political economy. Macmillan, London
- Kaldor N (1955) Alternative theories of distribution. *Rev Econ Stud* 23(2):83–100
- Kaldor N (1957) A model of economic growth. *Econ J* 67(268):591–624
- Kalecki M (1971) Selected essays on the dynamics of the capitalist economy 1933–1970. Cambridge University Press, Cambridge
- Keen S (2014) Secular stagnation and endogenous money. *Real World Econ Rev* 66:2–11
- Keynes JM (1971) Collected writings, vol XIII. Macmillan, London
- Kondepudi D, Prigogine I (1998) Modern thermodynamics: from heat engines to dissipative structures. Wiley, Chichester
- Krugman P (2012) End this depression now! WW Norton & Company, New York
- Mankiw N (2011) Principles of macroeconomics, 6th edn. Worth Publishers, New York
- Marshall A (1890) Principles of economics. Macmillan, London
- Martinez-Alier J (2007) Marxism, social metabolism, and international trade. In: Hornborg A, McNeill JR, Martinez-Alier J (eds) Rethinking environmental history: world-system history and global environmental change, AltaMira Press, Plymouth, pp 221–237
- Marx K (1867) Capital, volume I. New Left Review, Harmondsworth



- McLeay M, Radia A, Thomas R (2014) Money creation in the modern economy. *Bank Engl Q Bull* 1:14–27
- McNeill JR (2000) *Something new under the Sun: an environmental history of the twentieth-century world*. W. W. Norton & Company, New York
- Menger C (1871) *Principles of economics*. University Press, New York
- Murphy D, Hall C (2011) Energy return on investment, peak oil, and the end of economic growth. *Ann NY Acad Sci* 1219(1):52–72
- Nitzan J, Bichler S (2009) *Capital as power: a study of order and creorder*. Routledge, New York
- Parguez A, Seccareccia M (2000) The credit theory of money: the monetary circuit approach. In: Smithin J (ed) *What is money?*, Routledge, New York, pp 101–23
- Pasinetti LL (1962) Rate of profit and income distribution in relation to the rate of economic growth. *Rev Econ Stud* 29(4):267–279
- Piketty T (2014) *Capital in the twenty-first century*. Harvard University Press, Cambridge
- Reich R (2013) Higher wages can save America's economy— and its democracy. URL [http://www.salon.com/2013/09/03/higher\\_wages\\_can\\_save\\_americas\\_economy\\_and\\_its\\_democracy\\_partner/](http://www.salon.com/2013/09/03/higher_wages_can_save_americas_economy_and_its_democracy_partner/). Accessed 4 Sep 2014
- Ricardo D (1817) *Principles of political economy and taxation*. G. Bell and Sons, London
- Smith A (1776) *An inquiry into the nature and causes of the wealth of nations*. A. and C. Black, Edinburgh
- Stiglitz J (2013) Inequality is holding back the recovery. <http://opinionator.blogs.nytimes.com/2013/01/19/inequality-is-holding-back-the-recovery/>. Accessed 4 Sep 2014
- Walras L (1896) *Eléments d'économie politique pure, ou, Théorie de la richesse sociale*. F. Rouge
- Warr B (2009) REXS database. <https://sites.google.com/site/benjaminwarr/the-economic-growth-engine/rexs-database+>. Accessed 9 April 2012
- Wicksteed PH (1894) *An essay on the co-ordination of the laws of distribution* (1932 Edition). London School of Economics, London

## Chapter 4

# Institutional Size

*The Modern Corporation has undermined the preconceptions of classical economic theory as effectively as the quantum undermined classical physics at the beginning of the 20th century.*

– Gardiner Means (1957, p. 287)

The scale of both modern governments and modern corporations has no corollary in human history. In 1957, Gardiner Means observed: “We now have single corporate enterprises employing hundreds of thousands of workers, having hundreds of thousands of stockholders, using billions of dollars worth of the instruments of production serving millions of customers and controlled by a single management group” (p. 288). Despite the ubiquity of large corporations and governments, neoclassical growth theory assumes away their existence. Is such an approach empirically justified? This chapter aims to find out.

The layout of this chapter is as follows: I first dissect the neoclassical treatment of both the firm and government in order to understand why large institutions are systematically ignored. I then look at alternative theories of large institutions, before proceeding to an investigation of the empirical evidence linking the growth of energy consumption to the growth of both corporate and government employment concentration. Finally, I conclude by hypothesizing that the link between energy consumption and institutional size might be explained in terms of hierarchy.

### 4.1 Institutions in Neoclassical Growth Theory

The two main institutions that dominate the modern landscape are governments and corporations (or in economic jargon, *firms*). The importance of the firm is well recognized in neoclassical theory—indeed, it is the principle unit of production. However, because of a series of assumptions needed to justify basic neoclassical tenets, *firm size* plays no role in neoclassical growth theory. Government, on the other hand, is exogenous to neoclassical theory itself—a theory that is primarily aimed at modelling the undisturbed free market. Not surprisingly, government is completely absent from the neoclassical Solow–Swan model. In order to understand

this neglect of large institutions, we need to better understand the basic postulates of neoclassical theory. We begin with the model of the firm.<sup>1</sup>

While the firm is the basic unit of production, little is said of its actual composition. Instead, it is treated as a *black box*—inputs and outputs are known, but the internal workings of the firm remain sealed from view. All that is known about its behavior is that it maximizes profits. Each firm is assumed to have a production function that allows a mathematical mapping of inputs onto outputs. Neoclassical growth theory then assumes that all such firms can be modelled by a single, aggregate production function. It turns out that this approach requires two basic assumptions: (1) all firms (and the economy as a whole) experience constant returns to scale; and (2) the economy is perfectly competitive.

Constant returns to scale is a property of a production function  $Y(K, L)$  such that increases in the scale of capital and labor inputs yield a corresponding increase in total output. Stated mathematically, this becomes:<sup>2</sup>

$$Y(cK, cL) = cY(K, L) \quad (4.1)$$

While, in principle, the production function of individual firms can have either constant, increasing, or decreasing returns to scale, in order to maintain compatibility with the marginal productivity theory of distribution, one must assume constant returns to scale. When formulating marginal productivity distribution theory, neoclassical economists ran into an “adding up” problem. Joan Robinson summarizes this issue: “How do we know that, if each factor is paid its marginal product, the total product is disposed of without residue, positive or negative?” (1934, p. 398).

Wicksteed (1894) formulated an elegant solution to this problem using a theorem developed by Euler. He showed that if one assumes that production exhibits constant returns to scale, then Euler’s theorem could be used to “prove” that each factor receives payment in exact accordance with its marginal productivity, thus solving the adding up problem.<sup>3</sup> As a result of this theorem, neoclassical growth theory (which maintains compatibility with neoclassical distribution theory) is forced to assume that all firms have constant returns to scale. This means that size is neither an advantage nor disadvantage: all firms, large or small, are given the same production function.

---

<sup>1</sup> To be fair, numerous scholars have attempted to add government to neoclassical growth theory (for instance Carboni and Medda 2011). However, in most applications, government remains absent.

<sup>2</sup> Quite curiously, the production function does not include energy, the single factor one might think is most important in making something.

<sup>3</sup> As used by Wicksteed, Euler’s theorem states that if  $Y = f(a, b, c, \dots)$  is a production function with factors of production  $a, b, c, \dots$  that exhibits constant returns to scale, then  $Y = a \frac{\partial Y}{\partial a} + b \frac{\partial Y}{\partial b} + c \frac{\partial Y}{\partial c} + \dots$ . That is, output  $Y$  is guaranteed to be the sum of the quantity of each factor times its marginal productivity.

We now turn our attention to the assumption of perfect competition. Neoclassical theory predicts that under conditions of perfect competition, markets will allocate resources in the most efficient manner possible (given existing patterns of distribution).<sup>4</sup> Here, perfect competition is used specifically to mean that firms are *price-takers*: they have no market power that allows them to dictate the price of their output. However, as Coase (1937) noted, this results in a paradox. While the firm is the basic unit of production in neoclassical theory, the theorized efficiency of perfect competition implies that firms should *not exist*. According to neoclassical logic, the most efficient form of production should occur when competition is most atomistic—when every individual is self-employed.

Given this paradox, why does neoclassical theory continue to assume perfect competition? Steve Keen (2001) argues that it is because without it, the most basic tenet of neoclassical theory—the equilibrium-seeking price mechanism—cannot be justified. Before explaining the problem, let us first review the neoclassical explanation of the market. Arthur Salter says it best:

... the normal economic system works itself. For its current operation it is under no central control, it needs no central survey. Over the whole range of human activity and human need, supply is adjusted to demand, and production to consumption, by a process that is automatic, elastic, and responsive. (1921, p. 15)

In neoclassical theory, the equilibrium-seeking quality of the free market is explained by the forces of supply and demand. Market equilibrium occurs at the intersection of supply and demand curves. These theoretical curves, in turn, are explained in terms of the behavior of individual consumers and producers. The downward sloping demand curve is explained by the law of diminishing marginal utility: a consumer will derive a decreasing amount of pleasure from each additional unit of consumption. The upward sloping supply curve is explained by the law of increasing marginal costs: each additional unit of production is assumed to become costlier to produce.<sup>5</sup> As long as all firms are price-takers (meaning there is perfect competition), the resulting equilibrium quantity of production is such that the market price equals both the marginal utility of the buyer and the marginal cost of the supplier.

This elegant explanation of the price mechanism underlies all other aspects of neoclassical theory. Yet without perfect competition, it fails to function. If firms have even the slightest market power, then the equilibrium price will diverge from marginal costs, causing the theory to break down. Steve Keen summarizes: “Unless perfect competition rules, there is no supply curve” (2001, p. 101). Thus, the assumption of perfect competition is central to the internal consistency of neoclassical theory. When applied to neoclassical growth theory, this leads to the conclusion that the

---

<sup>4</sup> This is known as the first fundamental theorem of welfare economics: under conditions of perfect competition, market equilibrium is *Pareto-efficient*. It is impossible to make any one individual better off without making at least one individual worse off.

<sup>5</sup> One might protest that the reverse may actually be true—that each additional unit of production will cost less. While this may be true in reality, as Harold Lydall notes, “neoclassical theory is built on the ... assumption of absence of economies of scale” (1971 p. 91).

optimal growth path should be through atomistic competition: large firms should play no preferential role (indeed, they should not exist).

We now move on to the role of government in neoclassical theory. Here we find another paradox: the perfectly competitive free market is theorized to be completely self-regulating, meaning government intervention is redundant—or even harmful. Franklin Fisher notes that “the principal policy insight of economics [is] that a competitive price system produces desirable results and that government interference will generally lead to an inefficient allocation of resources” (1987, p. 26; cited in Murrell, 1991, p. 60). Yet markets cannot exist without the enforcement of a property rights regime—that is, without a large, coercive institution dedicated to enforcing market contracts. Milton Friedman summarizes: “The role of government ... is to do something that the market cannot do for itself, namely, to determine, arbitrate, and enforce the rules of the game” (1962, p. 27).

Thus, neoclassical theory explains the workings of a free market in the absence of government intervention; however, government intervention is a *prerequisite* for the existence of the free market. This paradox has never been resolved. However, the practical way forward is to acknowledge that government must exist, but to insist that it should be as small as possible. For instance, Friedman (ibid) insists that the proper role of government is solely to resolve market imperfections, namely monopoly.

Both in the case of firms and governments, neoclassical theory is unequivocal: smaller is better. Paul Walker (2007) notes that the absence of large institutions, in neoclassical theory, can best be understood by looking at the historical origins of the theory itself:

The interest in the price system, culminating in the ‘perfect competition’ model, has its intellectual origins in the eighteenth-century debate between free traders and mercantilists. This debate wasn’t about competition, in any meaningful sense, and it wasn’t about the existence and organisation of the firm; it was about the proper scope of government in an economy, and the model it gave rise to reflects this. The central question of the debate was, Is central planning necessary to avoid the problems of a chaotic economic system? Adam Smith famously answered no.

... The formal model that arose from this examination is one which abstracts completely from any form of centralised control in the economy. It is a model delineated by ‘perfect decentralisation’. Authority, be it in the form of a government or a firm or a household, plays no role in coordinating resources.

As neoclassical growth theory is primarily an extension of micro-foundations onto the macroscopic scale, it is forced to maintain this ambivalence toward large institutions. We can summarize this position as constituting the following implicit assumption:

**Institution Size Assumption:**

Large institutions are not important to growth.

Before testing this assumption, however, I review some of the more popular explanations for why large institutions exist, and how they influence growth.

## 4.2 Theories of Large Institutions

Throughout the course of the twentieth century, many theories have been produced that attempt to explain the growth of both firms and governments. For the sake of brevity, I focus here on only a few approaches that are most relevant to the present enquiry. We begin with theories of the firm.

### 4.2.1 Theories of the Firm

Perhaps the most popular theory of the firm is Ronald Coase's (1937) "transaction cost" approach. As discussed previously, Coase notes that the existence of firms represents a paradox for neoclassical theory: "If production is regulated by price movements, production could be carried on without any organization at all". Thus, he asks: "why is there any organization?" (p. 19) For Coase, the distinguishing characteristic of the firm is the *absence* of the price mechanism. He postulates that firms exist because there is a *cost* of using the market.

This simple postulate can then tell us why firms grow. Coase writes: "... a firm will tend to expand until the costs of organizing an extra transaction within the firm become equal to the costs of carrying out the same transaction by means of an exchange on the open market or the costs of organizing in another firm". Firm size is the simple result of an optimization between external *transaction* costs and internal organizational costs.

While the transaction cost theory of the firm is often heralded for its expansive explanatory scope, this is not necessarily indicative that the theory is sound. Karl Popper (1959) argued that the hallmark of a scientific theory lies not in its explanatory power, but in its ability to make falsifiable predictions. Nitzan and Bichler argue that Coasian theory is deficient in this regard:

The problem [with Coasian theory] is, first, that the cost of transactions (relative to not transacting) and the efficiency gains of transactions (relative to internalization) cannot be measured objectively; and, second, that it isn't even clear how to identify the relevant transactions in the first place. This measurement limbo makes marginal transaction costs ... unobservable; and with unobservable magnitudes, reality can never be at odds with the theory. ((2009), p. 340)

While the Coasian theory of the firm is popular among economists, an approach called the *resource-based view* has become popular in the field of strategic management (Barney 1991, Peteraf 1993). This theory proposes that firms can be thought of as a bundle of resources.<sup>6</sup> Specific resources owned by a firm can serve as a competitive advantage that can generate an above normal profit, causing the firm to grow. Thus,

---

<sup>6</sup> For the biophysical minded reader, note that since it is based in management studies, the resource-based view of the firm is mostly concerned with *human* resources.

the resource-based view argues that firm growth is the result of a resource-driven competitive advantage.

Edith Penrose (1955) proposed that it was the pool of top management talent that served as the most important scarce resource. She argued that, since management required in-depth knowledge of the internal workings of the firm, the pool of potential top managers could come only from within the firm's existing employment ranks. This has come to be known as the *managerial limits to growth* hypothesis, and implies that the "availability of top managerial and technical talent serves as the bottleneck for a firm's growth rate in a particular period of time" (Kor and Mahoney 2004, p. 184).

The validity of the resource-based theory of the firm (including the managerial limits to growth hypothesis) hinges on the proposed link between specific resources, competitive advantage and above normal profit. This implies a theory of value: the theory must stipulate exactly how a particular resource is transformed into a value-creating competitive advantage. However, Priem and Butler (2001) argue that the resource-based view advances a theory of value that is tautological. A firm's *value-add* is the result of a resource-driven competitive advantage. However, resources create this competitive advantage because they are (among other things) *valuable*.

While many volumes have been filled with writings on the nature of the firm, there is surprisingly little that is of use to the empirical researcher. Geroski summarizes the difficulties:

Very little in the theory of the firm as we know it is testable. Transaction cost based theories of the firm are usually driven by factors (such as asset specificity) which are difficult to observe, and, in any case, these theories *rarely make useful predictions* about the determinants of corporate growth rates. Recent resource-based theories of the firm have more to say about corporate growth, but the core competencies or internal assets/skills which they are based on are also *difficult to observe*. Even Penrose's famous 'managerial limitations to growth' hypothesis is based on coordination and team building costs which are *rarely directly measured* and never appear on corporate balance sheets. (1999, p. i)[emphasis added]

For those who think that theories should be empirically testable, current theories of the firm are quite unsatisfactory. The sensible conclusion is that we still have very little understanding of why firms grow.

We now move on to theories that connect firm size with economic growth. On this topic, the most influential thinker is undoubtedly Joseph Schumpeter (1942), who proposed that large firms are the engines of growth. This theory, which is diametrically opposed to the neoclassical perfect competition doctrine, is often called the *Schumpeterian Hypothesis*. Schumpeter proposed that growth occurred mainly because of *innovations*—the introduction of new products, technologies, and methods of production. However, this process required the purposeful allocation of resources toward innovation development. Large firms, he argued, could devote proportionally more resources toward innovation because their market power allowed them to earn

large monopoly profits. Thus, there should be positive returns to scale that lead to concentrated markets dominated by large firms.<sup>7</sup>

While straightforward conceptually, the link between innovation rates and firm size has proved remarkably difficult to test empirically. Syrneonidis (1996) reviews the body of empirical work on this issue and notes the dispersion of both null and significant results. Cohen and Levin note that the empirical results on this topic are “most accurately described as fragile” (1989, p. 1078). The main difficulty has to do with the ability to quantify “innovation”—an intrinsically fuzzy concept. Different metrics for innovation will yield wildly different results. For a good review of the many ways of quantifying innovative intensity, see de Jong (2000).

In my view, the search for a link between firm size and innovation rates amounts to a very narrow interpretation of the Schumpeterian hypothesis. In practice it will always be difficult to isolate the output of specific firms and to compare this with their relative size. Too many variables are involved, and information is too scarce. A more practical approach is to interpret the Schumpeterian hypothesis more broadly. Since large firms are theorized to be the engines of technological progress, it follows logically that increases in market concentration should go hand in hand with growth. Unlike the link between innovation rates and firm size, a hypothesized link between market concentration and growth is easily tested. In Sect. 4.3, I test this broadly interpreted Schumpeterian hypothesis.

### 4.2.2 *Theories of Government Size*

The prolonged growth in the size of the state has been the subject of intense academic debate for at least a century. In the following discussion, I first look at theories that explain government size. According to Holcombe (2005), these come in three main varieties: budget-maximization theories, rational-choice models, and path-dependent models. I then look at two theories that link government scale with growth: Baumol’s cost disease and Keynes’s stimulus theory.

We begin with budget-maximization theories, such as those proposed by Niskanen (1974) and Brennan and Buchanan (1980). These theories assume that government bureaucracies (or government as a whole) are self-serving entities that attempt to maximize their budgets, but are restrained by voters and/or an institutional framework such as the constitution. The viability of such an approach depends on the behavioral assumption that bureaucrats do indeed attempt to maximize their budget. While maximizing behavior is one of the fundamental postulates of neoclassical theory, if this is interpreted as the maximization of external payoffs (rather than internal utility), the evidence is quite clear: humans do *not* behave this way. In a remarkable study, Henrich et al. (2001) conduct a cross-cultural experiment in 16 small-scale

---

<sup>7</sup> Schumpeter also departed from neoclassical orthodoxy by asserting that large firms behaved co-respectively so as to avoid price competition.



societies using the Ultimatum Game<sup>8</sup>. They note that “the canonical model of the self-interested material payoff-maximizing actor is systematically violated” (2001, p. 77). If the average individual does not seek to maximize his payoff, then there is little reason to suspect that government bureaucrats would seek to maximize their budgets.

In rational-choice models (Meltzer and Richard 1981; Peltzman 1980), government is not a self-interested Leviathan; rather, it reflects the preferences of citizens as expressed through the democratic process. Holcombe (2005, p. 98) notes that “the implication is that government grows because people have a preference for larger government”. However, the general problem with this approach is that, absent an objective measure of an individual’s internal preferences, the theory becomes tautological: since government is assumed to reflect public preference, the growth of government must be the result of a preference for larger government.

The usual method for stipulating preferences is through a utility function, in which voters are treated as utility maximizing individuals who vote in their self-interest (for instance, (Meltzer and Richard 1981). However, Geoffrey Hodgson argues that this is problematic: utility maximization is unfalsifiable since “utility cannot be observed”. He continues, “utility maximization can be adapted to fit any form of behavior, including the behavior of nonhuman organisms” (2012, p. 94). For rational-choice practitioners, this is seen as a strength; from a Popperian standpoint, this infinite adaptability is indicative of pseudo-science.

The last class of models are called *path-dependent* because they theorize that government growth is the direct result of particular choices that are nonreversible. Perhaps the best example is the *ratchet hypothesis* first put forward by Peacock and Wiseman (1961). The theory is that governments grow in response to crises like wars and depressions. While there does seem to be good evidence that government grows during wars (especially in the USA), government expansion has been a ubiquitous phenomena in almost every country on the planet. It is hard to believe that all of this government expansion can be attributed to economic crisis and war, especially since the creation of large welfare states occurred mostly during the relatively peaceful post-WWII boom.

We now move on to the relation between government size and economic growth. Adolph Wagner (1911) hypothesized what has become known as *Wagner’s law of increasing state activity*: as the economy grows, the relative size of public expenditure will increase. The following question has occupied many economists’ minds: does the expansion of government *drive* economic growth, or does government grow *on*

---

<sup>8</sup> The Ultimatum Game involves the division of a sum of money between two people. One person initially has all the money. This person then makes an offer (to the other individual) that may be either accepted or rejected. In the former case, the offer proceeds, while in the latter case, both participants receive nothing. The canonical maximizing model predicts that the receiving individual should accept any offer, since something is better than nothing. Knowing this, the person making the offer should advance the smallest possible amount. Contrary to theory, (Henrich et al. 2001) find that initial offers were, on average, no smaller than 25 % of the total.

*the back* of the private economy? Keynesian theory argues for the former, while Baumol's cost disease argues for the latter.

In Keynesian theory, the basic idea is that government spending can be used to stimulate growth. However, this is seen more as a tool to dampen business cycles (and alleviate depressions) rather than as a tool for prolonged growth. The theory is straightforward: during an economic downturn, Keynes saw the problem as one of ineffective demand, which could be remedied by government deficit spending. Once the crisis ended, the idea was that government spending could decrease, allowing the repayment of public debt. While Keynesian theory asserts that government spending can stimulate growth, it says very little about why there has been a prolonged expansion in the size of government.

*Baumol's cost disease*, put forward by William Baumol (1967), reaches very different conclusions about the relation between government expansion and economic growth. Baumol made the distinction between the goods-producing sector and the service sector. The goods-producing sector, he argued, was characterized by rapidly increasing labor productivity. However, the service sector was marked by mostly stagnant labor productivity, due to its intrinsically labor-intensive nature (for instance, Baumol argued that the number of musicians needed to play a Beethoven symphony has not changed since the nineteenth century). Baumol argued that labor mobility would lead to uniform wage growth across all sectors. Thus, while wages in goods-producing sectors would remain tied to productivity, wages in the service sector would become decoupled from productivity. This is Baumol's cost disease: persistent inflation in the price of services.

Because government primarily provides services, it follows that it should be afflicted by this same cost disease. Thus a rise in the productivity of the goods-producing sector should lead to a larger proportion of the national income devoted to government. A key assumption of Baumol's theory is that the composition of output does not change with time (i.e., services and goods-production maintain a constant share of real value-added). Under this assumption, it follows that increased productivity in goods-producing sectors will lead to an increase in the employment of the service sector, including government.

The most important aspect of this argument is that this structural change should lead to stagnating growth. The logic is as follows: since productivity drives growth, but the service sector cannot really increase its productivity, increases in service employment should be detrimental to growth. Since government provides services, Baumol's cost disease leads to the conclusion that the scale of government should be negatively correlated with growth. While many countries are now experiencing stagnating growth, attributing this to Baumol's cost disease is problematic. For instance, US historical data shows that the service sector has been monotonically increasing in size for two centuries (see Chap. 5). If Baumol's cost disease were true, why is growth slowing only now? The main problem with Baumol's approach is that it hinges on a linear view of production. If services do not represent final output, but instead are an input to other sectors, Baumol's theory loses validity.

Thus, as with firms, it seems fair to conclude that the expansion of government, and its relation to economic growth, is poorly understood. In the next section, I

undertake an empirical analysis that differs from all previous studies in that it focuses on the relation between institutional size and the growth of *energy throughput*, not the growth of real GDP.

### 4.3 Institutional Size and the Growth of Energy Consumption: Empirical Evidence

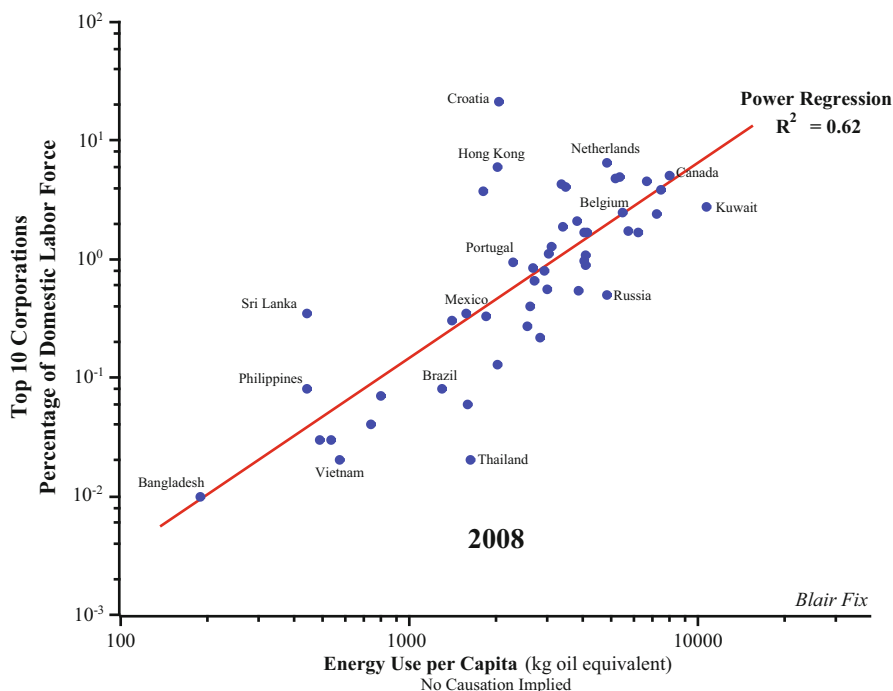
In order to proceed, we need metrics for both biophysical scale and institutional scale. As in previous chapters, I continue to use energy throughput as my indicator of the biophysical scale of the economy. However, because I engage in both international and subnational comparison (for which useful work data is unavailable) I use *primary energy* consumption per capita (or per labor hour, when available), rather than useful work.

There are also many possible measures of institutional scale. For instance Nitzan and Bichler (2009) investigate the scale of large corporations in terms of both profit and capitalization. However, this procedure does not work well when applied to the public sector. In order to use a methodology that is applicable to both public sector and private sector institutions, I use employment concentration (which is inspired by Nitzan and Bichler's concept of "breadth"). This approach has the advantage of using a biophysical unit (humans), whereas other measures of institutional size rely on monetary units.

When this methodology is applied to government, this means we express the employment of all levels of government as a fraction of total employment. When applied to firms, this means the fraction of total employment controlled by the top  $n$  firms (ranked by employment, where  $n$  is an arbitrary number chosen based on data availability). We then simply compare employment concentration and energy consumption series and look for correlation.

We begin with corporate employment concentration on the international scale. Figure 4.1 plots the national employment concentration of the top ten domestic firms against national energy consumption per capita. It reveals a rather tight correlation across more than three orders of magnitude of corporate concentration. Figure 4.2 moves to the national scale and plots time-series of the employment concentration of the top 200 US firms and national energy consumption per labor hour. Again, the correlation between the two series is fairly tight.

Notice that there is an abrupt change (from growth to contraction) in both time-series, circa 1970. Setting aside the real-world significance of this trend reversal, it has important *statistical* implications because it greatly lowers the possibility of spurious correlation. Gordon Tullock notes a problem with the least squares fitting procedure that can produce statistical artifacts: "any curve which is substantially smooth and reasonably monotonic has a high correlation with substantially any other curve meeting the same characteristics" (1983, p. 419). In a world in which many phenomena change monotonically over time, significant statistical correlation can occur between series that are, in reality, completely unrelated. However, the fact that

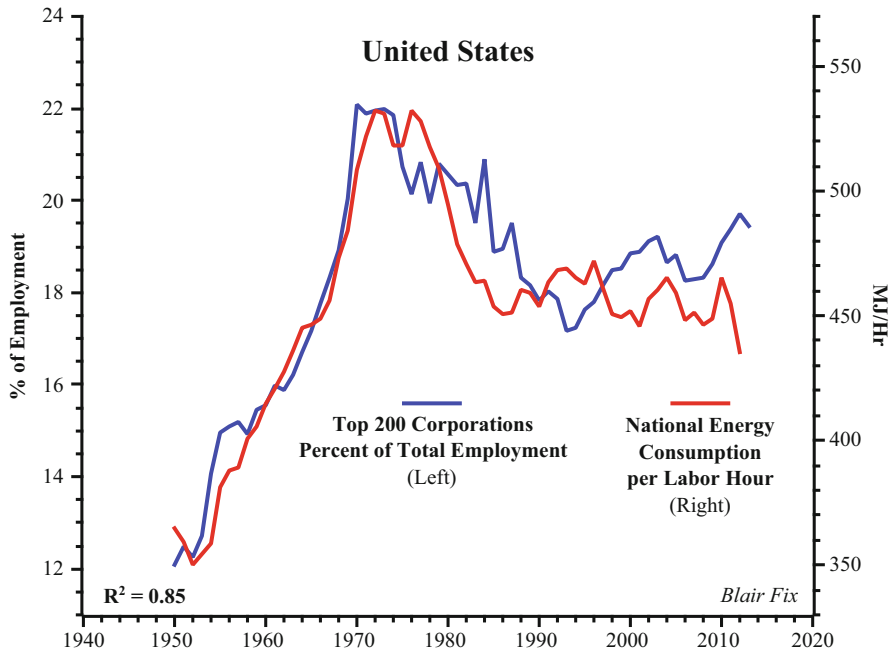


**Fig. 4.1** Domestic corporate employment concentration vs. energy use per capita. (Note: No causation is implied by the choice of x and y-axes variables. Sources: National energy use per capita and total labor force data is from the World Bank (indicator codes EG.USE.PCAP.KG.OE. and SL.TLF.TOTL.IN, respectively). Employment of top ten corporations (ranked by number of employees) is from COMPUSTAT Global Fundamentals (series EMP))

corporate concentration and energy consumption both rise *and* fall together, makes spurious correlation much less likely.

It could be argued that this trend reversal is the result of a change in the composition of the economy. Since the service sector is intrinsically less concentrated and less energy intensive than other sectors, a general shift towards this sector could conceivably produce a decline in energy consumption and employment concentration. However, the data plotted in Figs. 4.3 and 4.4 removes this possibility. Close to 1973 (the first energy crunch), both the service sector and the primary and secondary sectors experienced a similar trend reversal (here, primary and secondary sectors include industry and agriculture; the service sector includes all other sectors). Again, energy consumption and corporate employment concentration remain wellcorrelated.

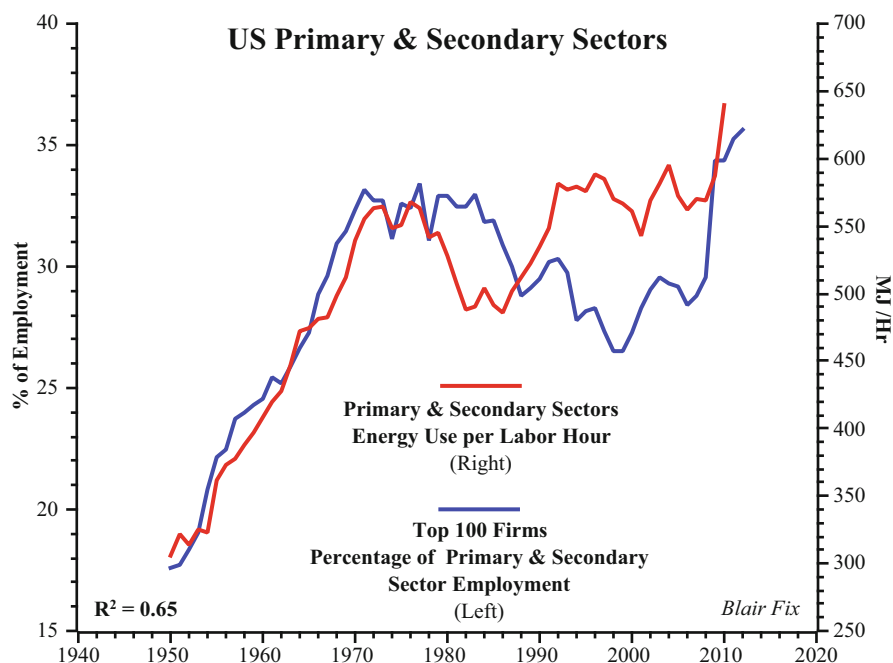
This evidence certainly does not bode well for the neoclassical assumption that large firms are not necessary for growth. On the other hand, it does support the Schumpeterian Hypthothesis that large corporations are the engines of growth. However, the problem with this latter conclusion is that we are measuring the throughput of the entire economy, not the productivity of large firms. Thus, we cannot necessarily



**Fig. 4.2** US Corporate employment concentration and energy use per capita. (Sources: Total US employment from Bureau of Economic Analysis (BEA) Tables 6.5 B-D (Full-time equivalent employees by Industry). Employment of top 200 corporations (ranked by number of employees) from COMPUSTAT (series DATA29). Total energy consumption from EIA Table 1.3 (Primary Energy Consumption by Source). Total labor hours from BEA Tables 6.9 B-D (Hours Worked by Full-Time and Part-Time Employees by Industry))

say that it is large firms that are driving growth (although this seems likely); it is conceivable that large firms are mostly growing by way of mergers and acquisitions—accumulating smaller, more productive firms during good times and divesting during bad times. While causation must remain ambiguous (for now), we *can* conclude that corporate employment concentration is inherently connected to the growth of energy consumption.

We now move on to the relation between government employment and energy consumption. We look first at the international scale. Figure 4.5 plots government employment concentration against domestic energy use per capita. The resulting trend is less clear-cut than for large firms, but is still present, especially if we remove the outlying “low four” nations from the regression (Moldova, Russia, Singapore, and South Korea). When we move to a time-series comparison within the United States (Fig. 4.6) there is also clear correlation. As with corporate employment concentration, government size plateaus at nearly the same time as energy consumption per labor hour. However, there is a large deviation during World War II, when the US government took command of the economy. This appears as a large pulse in employment concentration between 1941 and 1945. Interestingly, during the war,



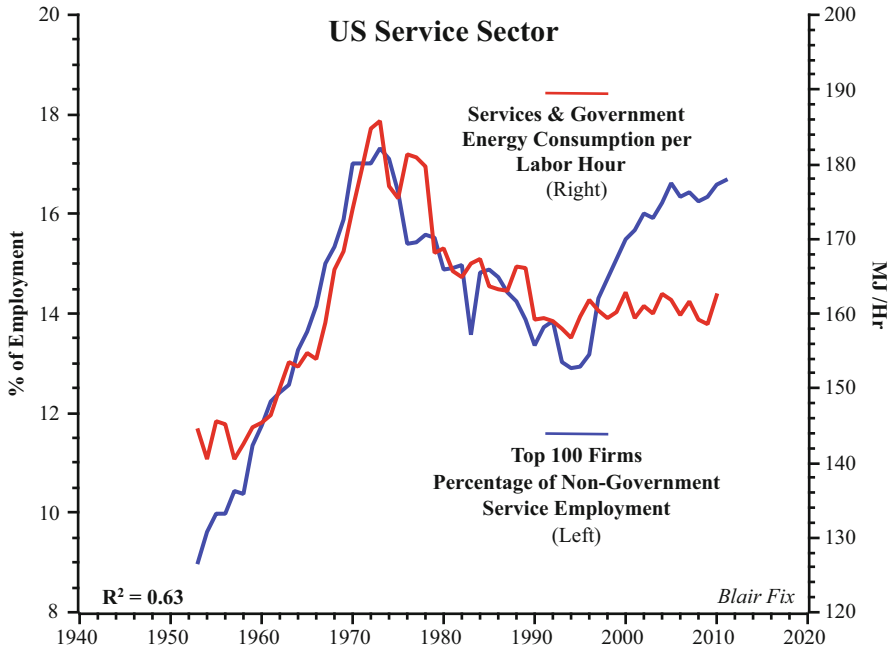
**Fig. 4.3** Primary and secondary sector employment concentration and sectoral energy use per labor hour (Sources: See Appendix)

energy use per labor hour actually declined. In keeping with the standard practice of many growth theorists, I do not include WWII in the regression.

In contradiction to neoclassical assumptions, the evidence suggests that the size of government is related to growth (at least, as growth is measured here). Indeed, the evidence is quite strong in suggesting that large institutions—both public and private—are an *integral* part of the biophysical growth process. Thus, we can conclude that the neoclassical assumptions that large institutions are unimportant for growth is *insidiously* untrue. But while the empirical evidence is clear, there is a distinct lack of theory that is able to explain it.

## 4.4 The Importance of Hierarchy

Over the last 10,000 years, humanity's transformation from a small-scale hunter-gatherer species to a globe-spanning industrial civilization has been nothing short of remarkable. What are the organizational mechanisms that facilitated this transformation? Economists typically argue that it was the price-mechanism (i.e., the "free market"). Prices, they argue, constitute a communication system that allows the spontaneous emergence of organization from otherwise unrelated activity.

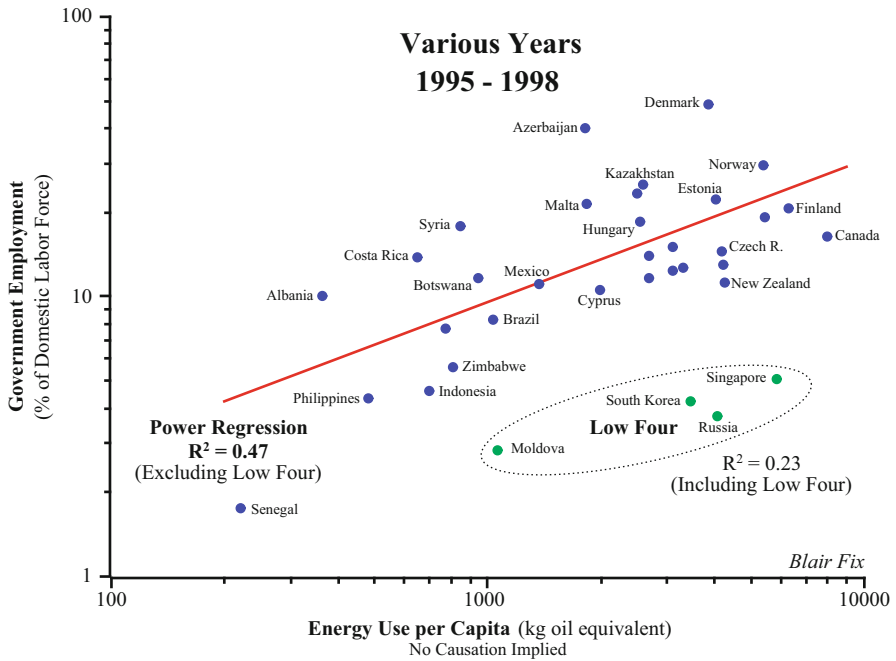


**Fig. 4.4** Service sector employment concentration and sectoral energy use per labor hour. (Sources: See Appendix.)

While I acknowledge the importance of prices, I pursue a different explanation. Anthropological evidence suggests that *hierarchy* was the most important organizational element for the large societies of antiquity. Using the empirical results from Sect. 4.3, I hypothesize that in the modern era, nothing much has changed: hierarchy (as embodied in large institutions) continues to form the backbone of society. Moreover, I hypothesize that increases in hierarchy are essential for growth.

I begin by noting that a species' ability to form social groups is, in large part, a function of genetic inheritance. Many animals (such as bears) are incapable of forming large groups, while others (such as wolves) do so naturally. All social organisms have evolved mechanisms that maintain the cohesiveness of their groups. It appears that in primates, this has involved the growth of large brains. In a remarkable study, anthropologist Robin Dunbar (1992) found that the group size of different primates was highly correlated with the relative size of their neocortex (compared to the rest of the brain). His conclusion was that neocortex size places an upper limit on the number of social relations that can be monitored by an individual. That is, brain size imposes a limit on group size. From his results on nonhuman primates, Dunbar (1993) extrapolated to find that human brain size predicted a group-size limit of about 150 (often called "Dunbar's number").

While this number should be considered exploratory, Dunbar notes that Neolithic villages had populations in this order of magnitude. Clearly, however, humans have



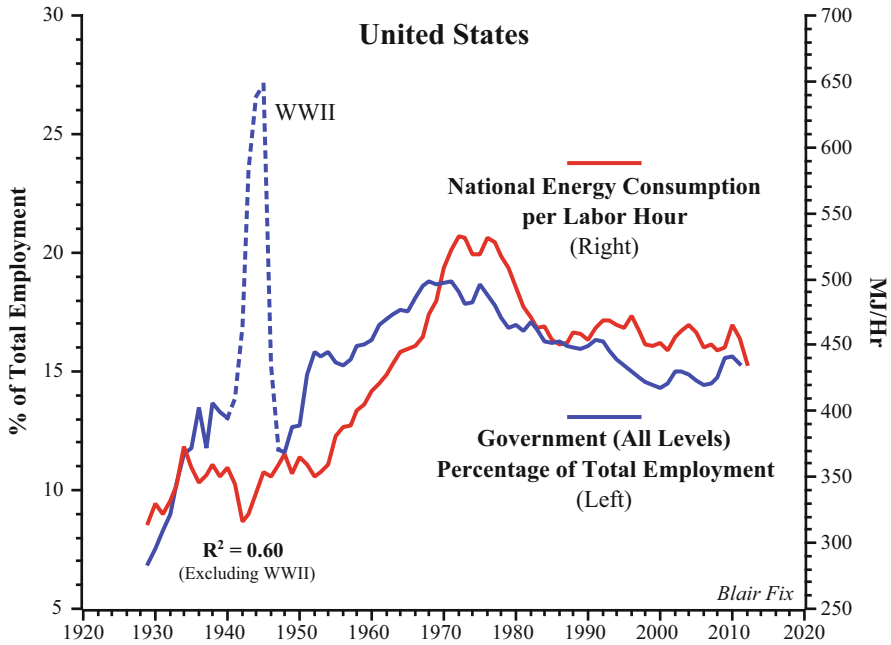
**Fig. 4.5** Government employment concentration vs. national energy use per capita. (Note: No causation is implied by the choice of x and y-axes variables. Sources: National energy use per capita and total labor force data is from the World Bank (indicator codes EG.USE.PCAP.KG.OE. and SL.TLF.TOTL.IN, respectively). Government employment is from Hammouya (1999), using total employment of the general government sector. Government employment data points vary between the years 1995–1998 and are matched with energy and labor force data for the appropriate year)

evolved ways of vastly exceeding this social scale: modern cities can surpass Dunbar’s number by five orders of magnitude. Moreover, it appears that social complexity increases in tandem with group size. In a landmark study of scaling effects within human society, George Zipf (1949) found that the diversity of business establishments scaled with city size according to a power law relation<sup>9</sup>. Among agrarian societies, Raoul Naroll (1956) found a scaling relation between the population of the largest settlement and the number of occupational specialities. Similarly, Robert Carneiro (1967) found that the organizational complexity of single-settlement societies (as measured by an index of 205 traits) scaled tightly with population.

As population increases, the evidence suggests that societies become more complex. How, then, does coordination and the maintenance of social cohesion occur? Dunbar argues that language is essential. However, while current research into the origin of language suggests that it is more than 100,000 years old (Perreault and

<sup>9</sup> A power law relation follows the formula  $y = ax^b$ , where  $a$  and  $b$  are constants.

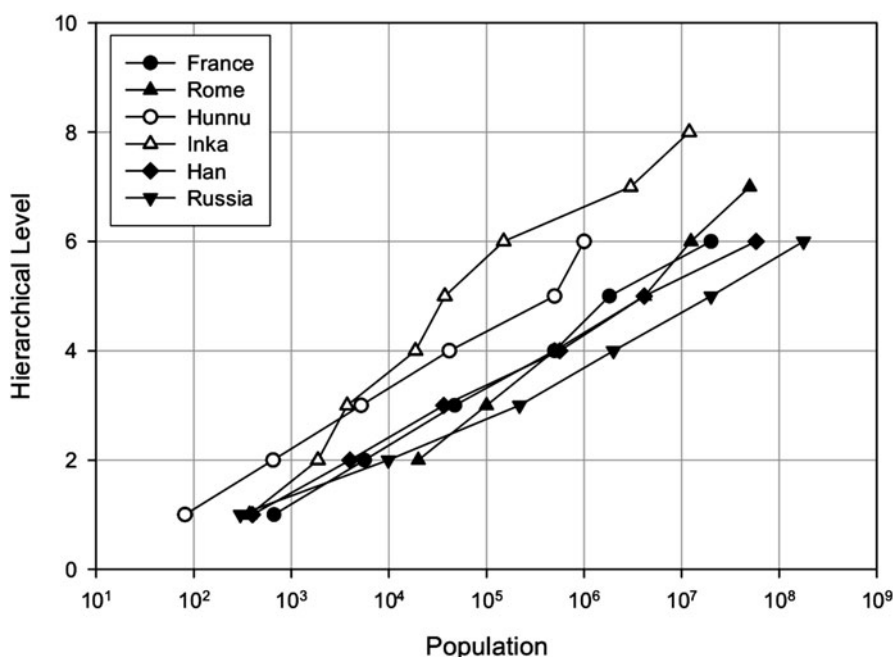




**Fig. 4.6** US government employment concentration and national energy use per capita. (Sources: US government and total employment data from BEA Tables 6.5 B-D, Full-Time Equivalent Employees by Industry. Total labor hours for 1948–2012 from BEA Tables 6.9 B-D. Data for 1929–1947 is from [BEA \(1973\)](#) Series A-68 (via Warr’s REXS database). US energy consumption from EIA Table 1.3, Primary Energy Consumption by Source. Energy data for 1929–1948 from REXS database Table 1.F.a, Exergy Inputs by Type. Exergy is converted to energy (heat content) using constants found in Ayres and Warr [2005](#), Tables A.1–A.6)

Mathew [2012](#)), large-scale human societies (i.e., civilizations) are generally regarded as about 5000 years old. John Gowdy and Lisi Krall note that after the transition to large-scale organization, “human society began to function like a single organism dedicated to the purpose of producing an economic surplus” ([2013](#), p. 138). The long delay between the origin of language and the transition to what Gowdy and Krall call “ultrasociality” suggests that some organizational strategy other than language has played a decisive role in increasing the complexity of human society.

Economists usually postulate that the market is responsible for modern economic organization—that it is the exchange-facilitating price mechanism that allows large-scale economies to spontaneously organize. While prices are undoubtedly important, they are, in many ways, an extension of language. If language allows a mapping of *words* onto an object, prices allow the mapping of an abstract *quantity*. To state another way, prices allow communication in the language of mathematics. Prices provide an ingenious mechanism for facilitating exchange, but they do not remove the necessity for direct interpersonal contact. So long as exchange is conducted solely by individuals, each person could conceivably transact with any other member of the

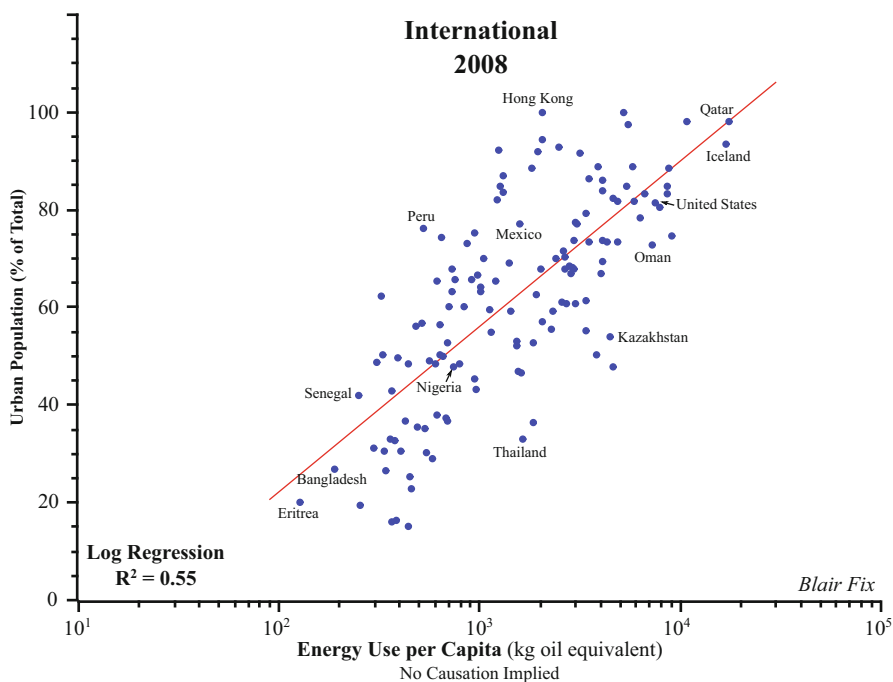


**Fig. 4.7** Hierarchical complexity (number of distinct administrative levels) vs. population of six historical empires. (Source: Turchin and Gavrillets 2009)

population, meaning an individual's possible interpersonal connections will scale with population size. However, any mechanism that allows group size to grow well beyond Dunbar's number must act to *limit* the number of interpersonal connections. This means that it is highly unlikely that prices alone could facilitate the extreme complexity of modern society.

Unlike prices, Turchin and Gavrillets (2009) note that hierarchical organization allows group size to grow without a corresponding increase in the number of interpersonal relationships. A member of a hierarchy needs to have a relationship only with his direct superior and his direct subordinates. If the number of subordinates is  $n$  (the *span of control*), then the number of direct interpersonal relationships required by any member is at most  $n + 1$ . Looking at historical agrarian empires, Turchin and Gavrillets find a strong scaling relation between population size and the number of administrative levels within the society (Fig. 4.7). Thus, evidence suggests that both hierarchy and social complexity scale with population. It seems warranted to conclude that, at least for agrarian societies, an increase in social complexity requires an increase in hierarchy.

But does this connection continue into the modern era? It is possible that modern growth has somehow changed this historical relation. As noted above, the anthropological literature has mostly focused on the relation between population size, social complexity, and hierarchy. However, my interest is in the connection between



**Fig. 4.8** Urbanization vs. energy use per capita. (Source: Data from World Bank Development Index, series SP.URB.TOTL.IN.ZS (urban population as portion of total) and EG.USE.PCAP.KG.OE (energy consumption per capita))

*energy consumption*, social complexity, and hierarchy. In order to make this connection, we begin by arguing that occupational specialization (social complexity) must intrinsically be accompanied by an increase in energy consumption.

As occupational specialization proceeds, new products are added but older products do not necessarily disappear. Indeed, in many cases, new products fuel demand for older ones. For instance, the introduction of the automobile vastly increased the demand for steel. If the demand for these older products does not fall, but the proportion of society engaged in making them does, it follows that labor productivity in these occupations must increase. This, in turn, requires an increase in the use of labor-augmenting technology. But, since it takes energy to make and/or fuel such technology, we can conclude that specialization can occur only through an increase in the per capita energy controlled by humans.

Ideally, we would be able to support this conceptual argument with direct empirical data linking energy consumption with social complexity. However, because I am not yet aware of data that would allow such a connection, we must make it via urbanization. As noted above, the level of specialization has been found to scale with the population of cities. Thus, it follows that levels of urbanization should be a strong indicator of social complexity. If we can establish a link between energy consumption and *urbanization*, then we can indirectly establish a link between energy consumption and *social complexity*. As Fig. 4.8 shows, urbanization is wellcorrelated with

energy consumption. From this evidence, we can make the preliminary assertion that as societies increase their energy consumption, they become more complex.

But in the industrial era, does this increase in complexity still require an increase in hierarchy as it did in the preindustrial era? I argue that it does. In economic theory (and most other social-scientific theory) public institutions (i.e., governments) are treated quite differently from private institutions (i.e., corporations). There are many good reasons for creating a dichotomy between the two, the most important of which is that the primary goal of many private institutions is to make a profit, while governments have no such clear-cut objective. Despite these obvious differences, large institutions, whether public or private, all have remarkably similar organizational structures. Both are bureaucratic institutions, characterized by a formalized chain of command. That is to say, they both make use of social *hierarchy*. Thus, I propose that the employment concentration of such large institutions is a good measure of the overall degree of hierarchy within a society. As Figs. 4.1–4.6 show, both corporate and government employment concentration scale with energy consumption.

My hypothesis, then, is that increases in energy consumption require an increase in social complexity, which requires an increase in hierarchy. This hierarchy, in turn, takes institutional form as either big government or big corporations. If this hypothesis is correct, it provides a powerful unifying force for social-scientific theory.

## 4.5 Conclusions: Size Matters

Recapping our findings, we discovered that energy consumption per capita—which I interpreted as an indicator of biophysical growth—is highly correlated with the employment concentration of large institutions. This runs directly counter to the neoclassical assumption that large institutions are unimportant for growth. Indeed, if the “market” of neoclassical theory is defined as exchange between individuals, we can state that biophysical growth involves the systematic replacement of “market” relationships with bureaucratic, hierarchical ones.

I argued that, while there is voluminous literature on the topic, our current understanding of large institutions is actually very limited. I’ve offered an explanation that attempts to make the link between energy consumption, social complexity, and hierarchy. I argue that, as complexity increases, coordination between humans must increase. However, this coordination must be accomplished without increasing the number of interpersonal relationships, which Dunbar argues is limited by the capacity of the human brain. Hierarchy is an essential tool for this task: it allows organizational size to increase while the number of required interpersonal relationships remains fixed.

Lund et al. (2007) articulate the importance of hierarchy, not only among humans, but throughout the animal kingdom:

Status, prestige, esteem, honor, respect, and rank are distributed differentially in all known groups. All human groups have social hierarchies, whether formal, as in the military or in business settings, or informal, as in cliques or other casual social groups.

... Social hierarchies are not unique to humans. Since the discovery of the pecking order among hens, the status hierarchy has been considered one of the main forms of social organization in animals. Indeed, status hierarchies are ubiquitous among the closest primate relatives of humans, the chimpanzee.

Despite its ubiquity in all known human societies, hierarchy is systematically ignored by neoclassical economics. As Nitzan and Bichler note, this has to do with the disciplinary split between “economics” and “politics”:

... from [Adam] Smith onward, it became increasingly customary to separate human actions into two distinct spheres, ‘vertical’ and ‘horizontal’. The vertical dimension revolves around power, authority, command, manipulation and dissonance. Academically, it belongs to the realm of politics. The horizontal axis centres around well-being, free choice, exchange and equilibrium - the academic preoccupation of economists. (2000, p. 67)

Thus, neoclassical growth theory explicitly assumes that growth is a purely “economic” phenomena that requires no centralization of power. However, all the evidence presented in this chapter suggests the opposite conclusion—namely, that growth is predicated on the increasing scope of large, hierarchical institutions.

This represents an interesting challenge to conventional economic theory, which treats centralized power as a “distortion” to its otherwise atomistic premise. Two future theoretical paths are possible: either a purely “economic” model of growth can be made to fit reality by alluding to frequent exogenous “shocks”, or a model can be developed in which hierarchy (and centralized power) plays an important role.

## References

- Ayres R, Warr B (2005) Accounting for growth: the role of physical work. *Struct Change Econ Dyn* 16(2):181–209
- Barney J (1991) Firm resources and sustained competitive advantage. *J Manage* 17(1):99–120
- Baumol WJ (1967) Macroeconomics of unbalanced growth: the anatomy of urban crisis. *Am Econ Rev* 57(3):415–426
- BEA (1973) Long term economic growth, 1860–1970. US Bureau of Economic Analysis, Washington, D.C.
- Brennan G, Buchanan JM (1980) The power to tax: analytic foundations of a fiscal constitution. Cambridge University Press, New York
- Carboni OA, Medda G (2011) Government spending and growth in a neoclassical model. *Math Financial Econ* 4(4):269–285
- Carneiro RL (1967) On the relationship between size of population and complexity of social organization. *Southwest J Anthropol* 23(3):234–243
- Coase RH (1937) The nature of the firm. *Economica* 4(16):386–405
- Cohen WM, Levin RC (1989) Empirical studies of innovation and market structure. In: Schmalensee R, Willig R (eds) *Handbook of industrial organization*, vol 2. Elsevier, New York, pp 1059–1107
- De Jong JPJ (2000) Measuring innovative intensity. EIM Business Policy Research Report 9912/A
- Dunbar RI (1992) Neocortex size as a constraint on group size in primates. *J Human Evol* 22(6):469–493
- Dunbar RI (1993) Coevolution of neocortical size, group size and language in humans. *Behav Brain Sci* 16(4):681–694

- Fisher F (1987) Adjustment processes and stability. In: Milgate M, Newman P (eds) *The new Palgrave: a dictionary of economics*, (vol 26–29). Palgrave Macmillan, London
- Friedman M (1962) *Capitalism and freedom*. University of Chicago Press, Chicago
- Geroski PA (1999) *The growth of firms in theory and in practice*. Center for Economic Policy Research 2092
- Gowdy J, Krall L (2013) The ultrasocial origin of the anthropocene. *Ecol Econ* 95:137–147
- Hammouya M (1999) Statistics on public sector employment: methodology, structures and trends. Working paper, International Labour Office, Geneva
- Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H, McElreath R (2001) In search of homo economicus: behavioral experiments in 15 small-scale societies. *Am Econ Rev* 91(2):73–78
- Hodgson GM (2012) On the limits of rational choice theory. *Econ Thought* 1(1):94–108. (History, Philosophy, and Methodology)
- Holcombe RG (2005) Government growth in the twenty-first century. *Public Choice* 124(1–2):95–114
- Keen S (2001) *Debunking economics: the naked emperor of the social sciences*. Zed Books, New York
- Kor YY, Mahoney JT (2004) Edith Penrose's (1959) contributions to the resource-based view of strategic management. *J Manage Stud* 41(1):183–191
- Lund OCH, Tamnes CK, Moestue C, Buss DM, Vollrath M (2007) Tactics of hierarchy negotiation. *J Res Pers* 41(1):25–44
- Lydall H (1971) A theory of distribution and growth with economies of scale. *Econ J* 81(321):91–112
- Means GC (1957) Collective capitalism and economic theory. *Science* 126(3268):287–293
- Meltzer AH, Richard SF (1981) A rational theory of the size of government. *J Polit Econ* 89(5):914–927
- Murrell P (1991) Can neoclassical economics underpin the reform of centrally planned economies? *J Econ Perspect* 5(4):59–76
- Naroll R (1956) A preliminary index of social development. *Am Anthropol* 58(4):687–715
- Niskanen WA (1974) *Bureaucracy and representative government*. Transaction Publishers, Piscataway
- Nitzan J, Bichler S (2000) Capital accumulation: breaking the dualism of 'economics' and 'politics'. In: Palan R (ed) *Global political economy: contemporary theories*. Routledge, New York
- Nitzan J, Bichler S (2009) *Capital as power: a study of order and creorder*. Routledge, New York
- Peacock AT, Wiseman J (1961) *The growth of government expenditures in the United Kingdom*. Princeton University Press, Princeton
- Peltzman S (1980) The growth of government. *J Law Econ* 23(2):209–287
- Penrose E (1955) Limits to the growth and size of firms. *Am Econ Rev* 45(2):531–543
- Perreault C, Mathew S (2012) Dating the origin of language using phonemic diversity. *PLoS one* 7(4):e35289
- Peteraf MA (1993) The cornerstones of competitive advantage: a resource-based view. *Strateg Manage J* 14(3):179–191
- Popper K (1959) *The logic of scientific discovery*. Hutchinson & Co., New York
- Priem RL, Butler JE (2001) Tautology in the resource-based view and the implications of externally determined resource value: further comments. *Acad Manage Rev* 26(1):57–66
- Robinson J (1934) Euler's theorem and the problem of distribution. *Econ J* 44(175):398–414
- Salter JA (1921) *Allied shipping control*. Oxford: At the Clarendon Press, New York
- Schumpeter JA (1942) *Socialism, capitalism and democracy*. Harper and Brothers, New York
- Syrneonidis G (1996) Innovation, firm size and market structures: schumpeterian hypotheses and some new themes. OECD Economic Department Working Papers(161)
- Tullock G (1983) Further tests of a rational theory of the size of government. *Public Choice* 41(3):419–421
- Turchin P, Gavrilets S (2009) Evolution of complex hierarchical societies. *Social Evol Hist* 8(2):167–198

- Wagner A (1911) II. Staat in nationalökonomischer hinsicht. (The state in economic perspective).  
In: Handwörterbuch der Staatswissenschaften. Verlag Gustav Fischer, Jena
- Walker P (2007) The firm in neoclassical theory. <http://antidismal.blogspot.ca>. Accessed 21 July 2014
- Warr B (2009) REXS database. <https://sites.google.com/site/benjaminwarr/the-economic-growth-engine/rexs-database>. Accessed 9 April 2014
- Wicksteed PH (1894) An essay on the co-ordination of the laws of distribution (1932 Edition).  
London School of Economics, London
- Zipf GK (1949) Human behavior and the principle of least effort. Addison-Wesley, Cambridge

## Chapter 5

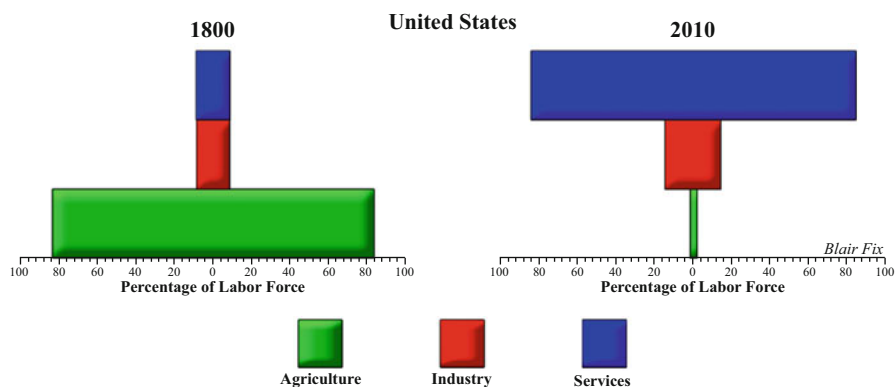
### Labor Structure

*The neoclassical view holds that sectoral composition is a relatively unimportant byproduct of growth.*

Cristina Echevarria (1997, p. 431)

Following the work of Allan Fisher (1935) and Colin Clark (1940), it has become common to classify labor into three sectors: agriculture, industry, and services. Since industrialization began, the change in the relative sizes of these sectors has been nothing short of remarkable. As Fig. 5.1 shows, between 1800 and 2010, the USA underwent a complete demographic *inversion*.

Despite the ubiquity of similar transformations among other industrialized nations, the canonical neoclassical model leaves no room for changes in labor demographics; instead, labor is modeled as an undifferentiated mass. This chapter investigates the empirical validity of such an approach. After investigating why neoclassical growth theory ignores labor demographics, I briefly review theories of structural change. I then develop a simple biophysical model of structural change based on three simple hypotheses. Finally, I look at the empirical relation between



**Fig. 5.1** A demographic inversion. (Sources: See Figure 5.10)



labor composition and energy consumption. Unlike the assumption made by neo-classical growth theory, the evidence suggests that labor structure is fundamentally connected to the growth process.

## 5.1 Labor in Neoclassical Growth Theory

Karl Whelan observes that “since the 1950s, the Solow–Ramsey<sup>1</sup> model of economic growth, which models all output in the economy in terms of a single aggregate production function, has been the canonical model of how the macroeconomy evolves in the long run” (2000, p. 1). By adopting a single sector model in which labor is treated as a homogeneous entity, the following assumption is implicit:

**Labor Structure Assumption:**

Changes in labor structure are not important for growth.

To be clear, in this chapter I test the assumptions made by the one-sector neoclassical *model*, not the opinions or assumptions of those who *apply* the model. The Solow–Swan model is explicit that only the *quantity* of labor input matters. However, some applied economists have attempted to adjust this labor quantity by accounting for changes in labor *quality*. For instance Denison (1962) and Schwartzman (1968) adjust the quantity of labor to account for changes in education. Perry (1971) adjusts for changes in the age–sex composition of the labor force, while Chinloy (1980) measures how sex, age, class, occupation, and education contribute to labor quality.

Unfortunately, measuring the quality of labor requires making inherently subjective decisions (as do all measures of quality). In the literature referenced above, a worker’s labor quality is defined in terms of his relative *productivity*. But herein lies the problem: the canonical approach is to use relative *wages* as a proxy for relative productivity (Perry et al. 1971). By doing so, researchers are explicitly assuming that neoclassical distribution theory is true—that “wage rates should measure marginal productivities” (Baily 1981, p. 8). Thus, by testing neoclassical growth theory using quality-adjusted labor inputs, one uses metrics that are circularly dependent on the theory being tested.

Furthermore, it is not clear that adjusting for differences in labor quality is admissible if the internal consistency of neoclassical theory is to be maintained. Aggregate production functions are supposed to represent the sum of all firm-specific production functions. However, Felipe and Fisher note that such aggregation is only theoretically permissible if one assumes that labor is homogeneous:

In the simplest case of constant returns, a labor aggregate will exist if and only if a given set of relative wages induces all firms to employ different labor in the same proportion ... The implication of these conditions is that the existence of a labor aggregate requires the *absence of specialization in employment*. . . (2003, p. 236) [emphasis added]

---

<sup>1</sup> The Solow–Ramsey model is similar to the Solow–Swan model. The former modifies the latter by introducing consumer optimization (i.e., utility functions).

By undertaking labor quality adjustments, one implicitly assumes that specialization of labor is important. However, Felipe and Fisher's aggregation conditions permit no specialization of labor. Thus, if neoclassical growth theory is to remain internally consistent, labor quality adjustments should not be used.

I should also note that while the one-sector model is canonical, some neoclassical economists have used multi-sector models to explain changes in labor structure. For instance, economists such as Echevarria (1997), Whelan (2003), Ngai and Pissarides (2004), and Foellmi and Zweimller (2008) have all used such multi-sector models to investigate structural change. Thus, the assumption that labor structure is unimportant does not apply to these models.

In fact, the empirical results of such models are often quite good. However, the ability to fit data, while important, is not the only criteria upon which a model must be judged. As Wassily Leontief notes, "what is really needed, in most cases, is a very difficult and seldom very neat assessment and verification of these assumptions in terms of observed facts (1971, p. 2)". The problem with the multi-sector models discussed above is that they begin with assumptions that, while plausible, are untestable.

There are two basic approaches to modeling structural change. The first approach is to endow the population with a utility function in which the preference for agricultural, industrial, and service commodities changes as a function of income (Echevarria 1997). The problem with this approach, however, is that utility functions cannot be observed, and are, therefore, unfalsifiable. Because of this, empirical success should be *expected*, since, as Hodgson notes, "we know in advance that any behavior can fit [utilitarian] theory" (2012, p. 104). The second approach is to assume differences in technological progress between sectors (for instance, Ngai and Pissarides 2004). However, since the technological progress term cannot be measured independently of the neoclassical model, this approach is also unfalsifiable.

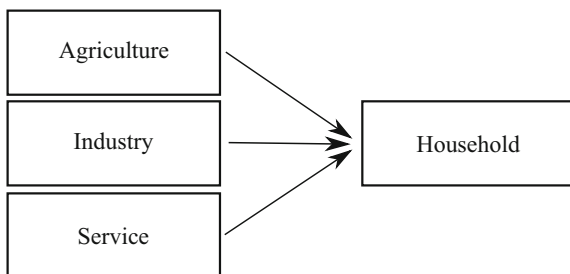
Thus, there exists considerable heterogeneity among neoclassical economists on the topic of structural change. However, to the extent that the one-sector model remains dominant, it seems fair to state that labor structure change plays no part in neoclassical theory.

## 5.2 Theories of Labor Structure Change

Many schools of labor transformation exist. For brevity, I review only three. In order of historical precedent, these are: the demand-side, the supply-side, and the metabolic approach. I review the basic tenets of each theory below.

We begin with the demand-side approach, which is rooted in the hypothesis of a hierarchy of human needs. In the mid-nineteenth century, Ernst Engel (1857) noticed that the portion of income spent on food tended to fall as income rose. He surmised that when the desire for food is satiated, individuals move on to more sophisticated needs. In the twentieth century, Alan Fisher (1935) and Colin Clark (1940) extended this hypothesis to the industrial sector. They argued that the desire for consumer goods

**Fig. 5.2** Linear treatment of the economy

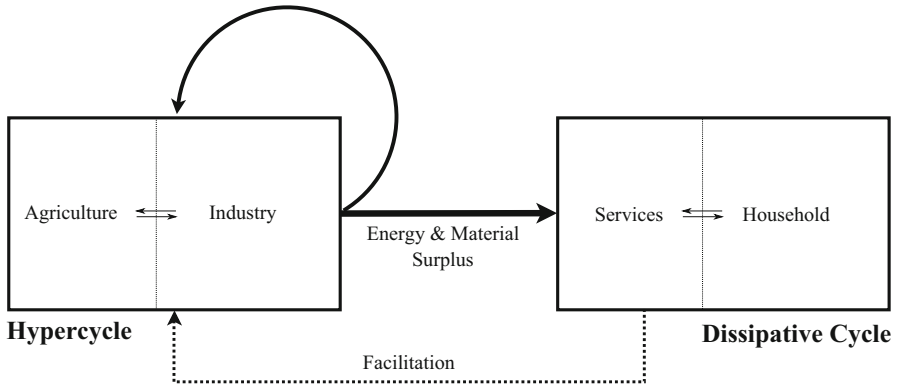


would become satiated, at which point people would switch to a desire for services. In economic jargon, the basic hypothesis of the demand-side approach is that the “income elasticity of demand” for goods is less than one (meaning the proportion of income spent on goods falls as income rises), while the income elasticity of demand for services is greater than one (meaning the proportion of income spent on services increases as income rises). Thus, the portion of income devoted to services tends to grow with income. As demand changes, so too should the sectoral composition of employment.

In contrast, the supply-side approach, advanced by William Baumol (1967), focuses on differences in the growth of labor productivity between sectors. While the goods-producing sector rapidly increases its labor productivity, the service sector—due to its inherently labor-intensive nature—increases its labor productivity much more slowly. Baumol *assumed* that the structure of final output remains constant (i.e., the inter-sectoral ratio of “real” output is fixed). Under this assumption, it follows that increase in the productivity of the goods-producing sector will lead to a decline in its employment share, while lagging productivity growth will yield an increase in service employment share. A major prediction of the supply-side approach is that the swelling of the low-productivity service sector should cause economic growth to slow. However, historical data show that the US service sector has been growing monotonically for more than two centuries (Fig. 5.7). Given this historical expansion, Baumol’s approach must explain why it is only recently that growth has slowed.

A shortcoming of both supply and demand-side approaches is that they adopt a simplistic, linear representation of the economy (Fig. 5.2). In both cases, only *final* consumption by households is considered. The problem, however, is that actual consumption patterns are much more complex: each sector consumes a portion of both its *own* output and the output of *all other* sectors. This means that a major fraction of the output of each sector is not destined for final consumption by households, but rather for inter-sectoral consumption. Thus, focusing only on final demand is misleading. Within supply-side literature, this realization has led to a refocus on *producer services*—services aimed at business rather than consumers (Greenfield 1966; Francois 1990; Greenhalgh and Gregory 2001).

Another shortcoming of both supply and demand-side approaches is that they focus only on production and ignore the importance of *consumption*. The metabolic approach advanced by Giampietro et al. (2012) attempts to correct this deficiency.



**Fig. 5.3** Hypercyclic and dissipative compartments of society. (Modeled after Giampietro et al. (2012). The hypercycle produces all the material needs of society. The material surplus from the hypercycle is dissipated by the dissipative cycle. In return, the dissipative cycle provides facilitation activity that is required for the proper functioning of the hypercycle)

Giampietro et al. credit George Zipf (1941) for first emphasizing the importance of consumption capacity:

“Zipf proposed a basic principle of socio-economic development: if an economy wants to be able to produce more, it has to invest more in consuming. This principle implies that socio-economic development must be based on achieving an internal balance between parallel investments both of human activity and of energy over the two compartments of production and consumption . . . ” (Giampietro et al. 2012, p. 157)

Building on the work of Eigen (1971) and Ulanowicz (1986), Giampietro et al. divide the economy into two compartments: the *hypercycle* and the *dissipative cycle* (Fig. 5.3). The hypercycle produces all of the material goods used by a society (including energy). The “hyper” portion of “hypercycle” refers to the fact that this sector produces a *surplus* of material goods. Thus, the energy sector produces more energy than it requires for its own purposes. If it did not, all of society would be engaged in energy harvesting activities.

This energetic surplus is used by all other sectors, including sectors within the hypercycle. The dissipative cycle, on the other hand, does not produce any material goods. Instead, it dissipates (a fancy word for “consumes”) the surplus of the hypercycle. In return, the dissipative cycle provides crucial *facilitation* activity to the hypercycle. A basic hypothesis of this approach is that such facilitation activity is essential to the proper functioning of the hypercycle. The important insight of Giampietro et al. is to place the service sector *inside* the dissipative cycle. Thus, the role of the service sector (and the household) is to both facilitate the activities of the hypercycle and consume its surplus.<sup>2</sup>

<sup>2</sup> The distinction between productive and unproductive labor is a long-running feature of Marxian political economy. For instance, Baran and Sweezy (1966) treat the service sector as unproductive,

The other main insight of the metabolic approach is its focus on the internal consumption of the hypercycle. In general, labor productivity gains will require greater amounts of hypercyclic energy and resource consumption. Thus, as human activity is transferred to the dissipative cycle, we can expect energy consumption to increase.

In the following sections, I develop a biophysical model of structural change that builds on the metabolic approach proposed by Giampietro et al.

### 5.3 Three Hypotheses

I contend that changes in labor composition can be explained using the following three hypotheses that build on the work of Giampietro et al. (2012; 2013), Cleveland et al. (1984), Hall et al. (1986), and other biophysical scholars who have insisted on placing energy at the center of analysis. Below, I explain the rationale behind each hypothesis and present supporting evidence.

1. All sectors attempt to increase their labor productivity.

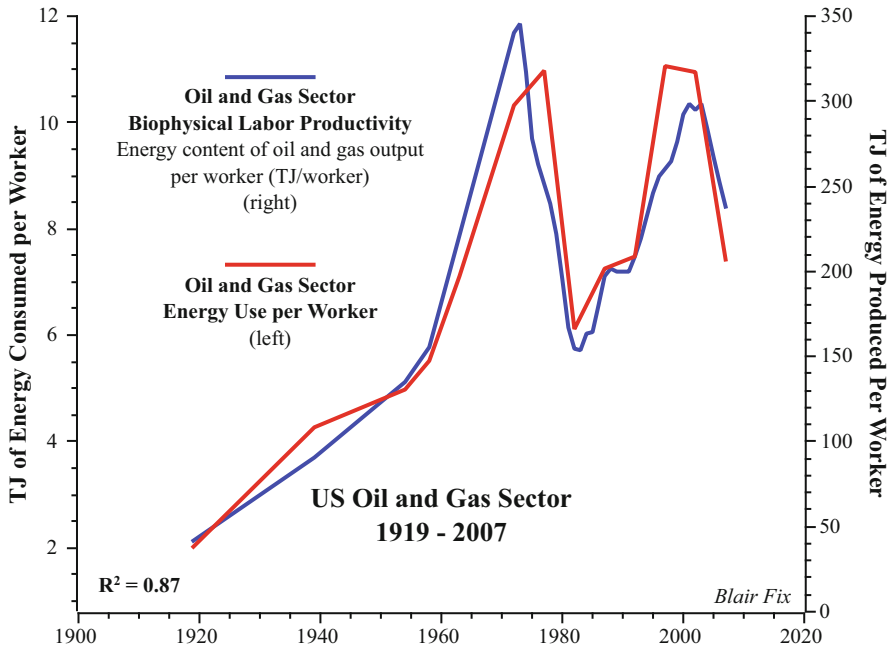
Among economists there is virtually unanimous recognition that the cost reduction imperative prompts firms to decrease their labor costs by increasing labor productivity. It is important to note that this impetus exists only so long as machine work is cheaper than employing human labor. This has been the case historically because exosomatic energy (primarily fossil fuel) was cheap. It may not be the case in the future.

2. Productivity increases require an increase in energy consumption.

This follows from thermodynamic principles: increases in labor productivity are accomplished primarily by substituting exosomatic for endosomatic work (nonhuman for human). This implies that labor productivity and energy consumption are fundamentally connected. For most sectors, this concept is difficult to test. Because sectoral output is generally heterogeneous and changes qualitatively with time, the quantity of output is usually calculated by deflating the nominal *value* of output by an average price index. However, this approach is subject to all the price index problems discussed in Sect 2.4. What we want is a physical measure of output that is not derived from prices. Fortunately the oil and gas sector is amenable to such an approach because its output is energy. Figure 5.4 plots the physical labor productivity of the oil and gas sector (energy output per worker) relative to its energy use per worker. As expected, the two series are tightly coupled, lending support to Hypothesis 2.

---

and hence, dissipative. For the present discussion, I do not use the Marxian language of “unproductive” because it implies that this labor is unnecessary. As Giampietro et al. note, dissipative activity is equally as important as hypercyclic activity.



**Fig. 5.4** Biophysical labor productivity and energy use per worker in the Oil & Gas sector. (Sources: US oil and gas production from HSUS Table Db157–158 (1919–1948) and EIA Table 1.3 (1949–2007). Direct energy consumption for the oil and gas sector from Guilford et al. (2011), Table 6. Oil and gas employment from 1958 and 1972 Census of Mineral Industries (1919–1972). Oil and gas employment is calculated indirectly as the difference between total employment of mineral industries and employment *excluding* oil and gas industries. Employment from 1972 to 2007 from BLS series CES1021100001 (1972–2007))

Correlation between energy consumption and productivity has also been found by Cleveland et al. (1984) and by Hall et al. (1986).

### 3. Increases in energy consumption require an increase in facilitation activity.

Here I use Douglass North’s (1990) proposed distinction between *transformation* activity and *transaction* activity (a distinction adopted by Giampietro et al. 2012). However, I refer to the latter category as *facilitation* activity, since I wish to disassociate from Coasian transaction cost theory. *Transformation* activities are those that are directly involved with manipulating material and energy, while *facilitation* activities are those that are indirectly involved, but still crucial for the production process as a whole. In general, agriculture and industry are classified as transformation activities, while service and household work is considered facilitation.

Increasing energy consumption (in order to increase productivity) does not occur through a simple scaling of the existing state of affairs. As I attempted to demonstrate

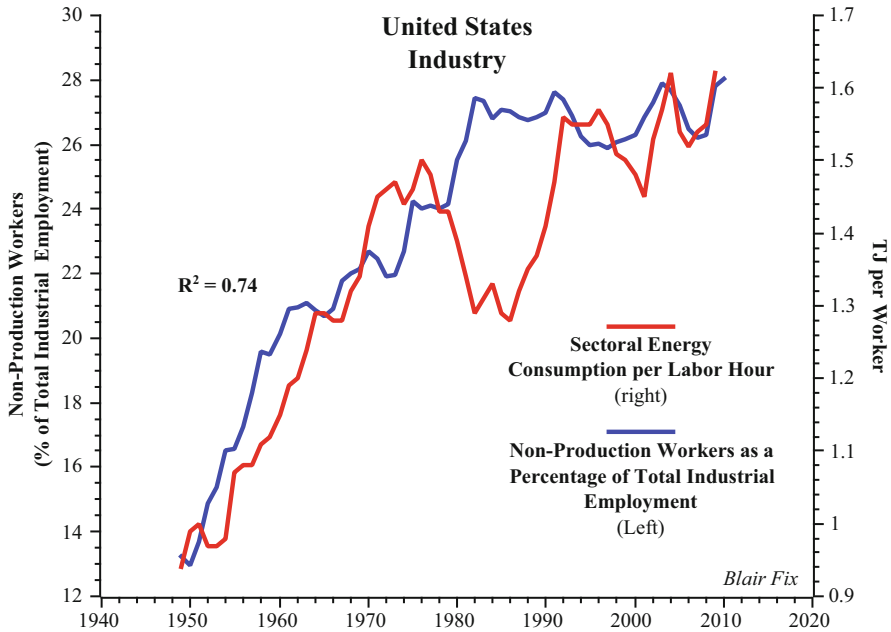
in Sect. 4.4, the growth of energy consumption entails an increase in the *complexity* of the economic system. Alongside this complexity comes an increasing need for occupational specialization. My hypothesis is that as energy consumption expands, facilitation activities must expand *faster* than transformation activities.

For an example of this process, take productivity increases in agriculture that are due to the increased use of tractors. Such productivity increases will be accompanied by an increase in agricultural energy consumption per worker, because tractors require energy to function. Since workers are required to *build* the tractors, industrial employment must also expand. Furthermore, a myriad of facilitation activities are required to support the use of tractors. For instance, mechanics will be required to service the tractors. Truck drivers will be required to transport both finished tractors and all of the ancillary products required to make the tractors. Engineers will be required to design the tractors. Educators will be required to train the skilled labor involved in designing, making, and servicing tractors. Government may be required to regulate the tractor industry. Banks are required to finance both the purchase of tractors by farmers, and the purchase of plant equipment by tractor manufacturers. The list of ancillary activity required to augment agricultural productivity is immense.

For the moment, we leave aside the question of *why* energy consumption increases might require a relative rise in facilitation activity (I do not have a satisfactory answer to this question). The more pertinent question is: how can we *measure* such activity without resorting to a “network tautology”? Note that for any given service sector activity, one can *always* demonstrate that this activity is “required” for hypercyclic productivity growth. Taken to the point of the absurd, suppose we have a hairdresser who cuts the hair of a massage therapist, who gives a massage to an educator, who trains a secretary, who works for an engineer who designs tractors. We can say, without too much of a stretch, that this hairdresser is *required* if tractors are to be used in agriculture. I defy anyone to find a facilitation activity that cannot somehow be connected to a hypercyclic activity. The problem is that this is true by *definition* of the fact that modern societies are well-connected networks in which almost no one lives in isolation. Thus any individual (or process) can easily be linked to any other individual (or process), so long as we allow enough degrees of separation.

By definition of their mutual existence in a networked society, most of the service sector is *required* for the hypercycle to function, and vice versa. But this is tautologically true, and therefore unsatisfactory. In order to test Hypothesis 3 (increases in energy consumption require an increase in facilitation activity) in a non-tautological manner, we must *restrict* the degree of separation between facilitation and transformation activity. As the degree of separation grows, we will asymptotically approach the network tautology. Therefore, the smaller the degree of separation, the better. It follows that the best possible test of Hypothesis 3 is to restrict analysis to one degree of separation—facilitation activity that has been *internalized* by the hypercycle.

The Bureau of Labor Statistics makes a distinction between “production” and “nonproduction” workers. The former are those who are directly involved with



**Fig. 5.5** Internal facilitation activity and energy use per worker in industry. (Sources: Note that “Industry” is taken to be equivalent to the BLS “goods-producing sector”. Total employment and production-worker employment of goods-producing sector from BLS (series CES0600000001 and CES0600000006 respectively). Industrial energy consumption data from EIA Table 2.1, Energy Consumption by Sector)

production (i.e. mechanics), while the latter are only indirectly involved (i.e. secretaries).<sup>3</sup> Thus, non-production employment represents the most direct form of facilitation activity. If Hypothesis 3 is true, we expect that increases in hypercyclic energy consumption (per worker) should lead to an increase in the proportion of nonproduction employment (in the hypercycle).

Since data on nonproduction employment is available only for nonfarm workers, we must restrict analysis to industry (instead of the entire hypercycle). Figure 5.5 plots the energy use per labor hour of US industry against the employment of non-production workers in industry (as a fraction of total industrial employment). The connection between the two series is clear: as we subsidize production workers with

<sup>3</sup> According to the BLS: “Production and related employees include working supervisors and all nonsupervisory employees (including group leaders and trainees) engaged in fabricating, processing, assembling, inspecting, receiving, storing, handling, packing, warehousing, shipping, trucking, hauling, maintenance, repair, janitorial, guard services, product development, auxiliary production for plant’s own use (for example, power plant), record-keeping, and other services closely associated with the above production operations” (BLS 2009, Ch. 2, p. 2). Nonproduction employees include everyone else.



more energy, we increase the productivity of each worker. This process increases the material surplus, which allows more ancillary workers. However, I argue that causation is not linear: increasing productivity both allows and *requires* more ancillary activity. Issues of causation aside, the results from Fig. 5.5 provide strong supporting evidence for Hypothesis 3.

## 5.4 Modelling Changes in Labor Structure

The three hypotheses outlined above can be used to create a simple, biophysical model of labor structure change. The model that I develop below uses only four exogenous, biophysical inputs: total population, agricultural productivity (proxied by wheat productivity), domestic fossil fuel output (or domestic energy consumption after 1949), and mining sector biophysical labor productivity (fossil fuel output per worker). All exogenous inputs come from historical data. Furthermore, the model contains only three parameters. Despite its simplicity, this model is able to reproduce historical trends with surprising accuracy.

I derive the model's basic equation from a mathematical identity: the labor ( $L$ ) required for any activity is equal to the quantity of output ( $Q$ ) times the labor per unit of output (note that even if we cannot measure output, this equation is true by definition):

$$L = Q \cdot \frac{L}{Q} \quad (5.1)$$

Since  $L/Q$  is the inverse of labor productivity ( $\rho$ ), we can rewrite Eq. 5.1 as:

$$L = \frac{Q}{\rho} \quad (5.2)$$

However, we want an equation for the *growth rate* of labor. Taking the derivative of Eq. 5.2 and expressing the result in terms of growth rates gives the following equation:<sup>4</sup>

$$\hat{L} = \hat{Q} - \hat{\rho} \quad (5.3)$$

This is the basic equation of “motion” for our model. It tells us that output growth represents a *centripetal* force that increases the growth of labor requirements. On the other hand, labor productivity growth represents a *centrifugal* force that decreases the growth of labor requirements. Given initial conditions (i.e. initial employment in each sector), Eq. 5.3 can be used to model changes in sectoral employment.

---

<sup>4</sup> Note that I use the convention of hat calculus where  $(\hat{\cdot})$  denotes growth rate. In hat calculus, the growth rate of a function ( $y$ ) is defined as the first derivative ( $y'$ ) divided by the original function:  $\hat{y} = y'/y$ . The product rule becomes a sum ( $\hat{x}\hat{y} = \hat{x} + \hat{y}$ ) and the quotient rule becomes a difference ( $\hat{\frac{x}{y}} = \hat{x} - \hat{y}$ ).

We begin with the agricultural sector. For simplicity, we assume that there is no net agricultural trade. Under this assumption, we can state that the demand for agricultural output is largely a function of total population, since per capita nutritional and caloric needs are mostly fixed. We ignore any changes in food preferences and/or total caloric intake that occurs with income, and we ignore any conversion of crops into biofuels.

Under these assumptions, the growth rate of agricultural output ( $\hat{Q}_{AG}$ ) is *constrained* by the historical growth rate of total population ( $\hat{P}_T$ ).<sup>5</sup> For simplicity, we assume that agricultural labor productivity is proportional to historical wheat productivity ( $\rho_{wheat}$ ), and that this proportionality is dictated by the constant  $\psi$ .<sup>6</sup> The resulting formula for the agricultural employment growth rate is given by Eq. 5.4. The value  $\psi = 0.71$  reproduces historical agricultural employment to an accuracy of  $R^2 = 0.95$  (see Fig. 5.7). If the model's assumptions are valid, this indicates that the productivity of the whole agricultural sector grew 71 % as quickly as wheat productivity.

$$\hat{L}_{AG} = \hat{P}_T - \psi \hat{\rho}_{wheat} \quad (5.4)$$

We now move on to industry. Again, industrial employment growth is given by the following equation:

$$\hat{L}_{IN} = \hat{Q}_{IN} - \hat{\rho}_{IN} \quad (5.5)$$

Recall that according to Hypothesis 2, labor productivity is proportional to energy use per worker. A corollary of this hypothesis is that *output* is proportional to energy consumption. We can show this by noting that labor productivity is simply output per worker. Thus, we can write Eq. 5.6. Since the denominators on the left and right side are equivalent, if Hypothesis 2 holds, we know that energy consumption is proportional to output. Another way of thinking about this is that output is *constrained* by energy consumption.

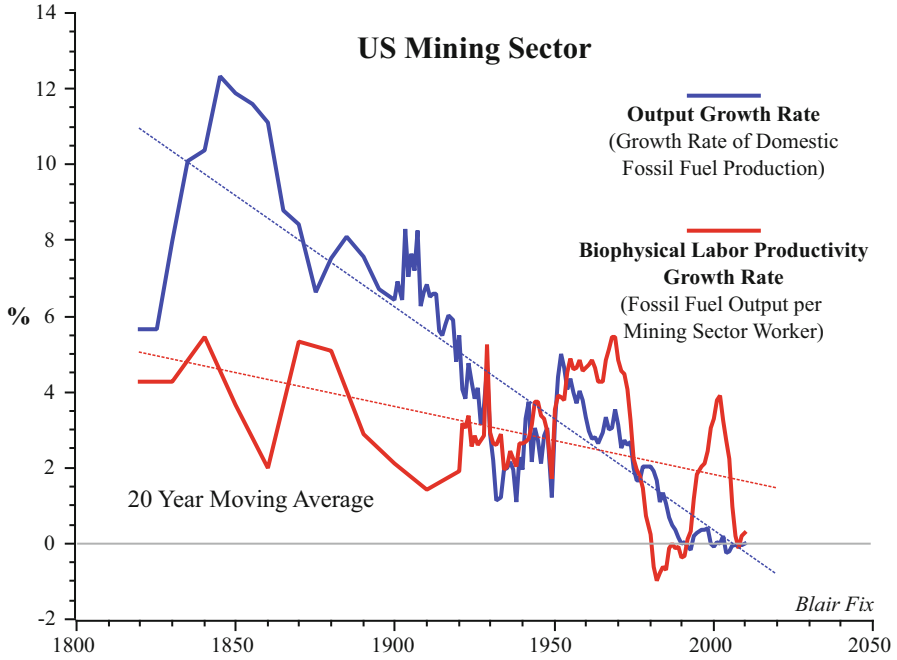
$$\frac{\text{output}}{\text{number of workers}} \propto \frac{\text{energy use}}{\text{number of workers}} \quad (5.6)$$

Thus, from Hypothesis 2, we assume that the output of industry is constrained by its ability to consume energy, and that this constraint is governed by the constant  $\sigma$ . Assuming a closed economy in which fossil fuel energy dominates, the energy available to industry is equal to the fossil fuel output of the mining sector ( $Q_M$ ) times the fraction of this output consumed by industry ( $f_{IN}$ ):

$$Q_{IN} = \sigma \cdot Q_M \cdot f_{IN} \quad (5.7)$$

<sup>5</sup> Because we are modeling agriculture employment, we take total population as given. When I say population “constrains” agricultural output, the reverse is also true—agricultural output growth constrains total population growth. Thus, we simply assume a coupled relation and leave aside issues of causation.

<sup>6</sup> Ideally we would link agricultural productivity back to agricultural energy consumption. However, I leave such an endeavor for future work.



**Fig. 5.6** Growth rates of output and labor productivity of the mining industry. (Sources: Fossil fuel production data from HSUS Table Db155 (1920–1948) & EIA Annual Energy Review 2011, Table 1.2 (1949–2011). Mining sector employment from HSUS Table Ba819 (1900–1910), HSUS Table Ba841 (1920–1957), & BLS CEU1021000001 (1958–2011))

We take the derivative of Eq. 5.7. Since  $\sigma$  is constant, it disappears, leaving only the following relation:

$$\hat{Q}_{IN} = \hat{Q}_M + \hat{f}_{IN} \quad (5.8)$$

This equation states that the industrial output growth rate is equal to the domestic fossil fuel output growth rate plus the rate of change of industry's share of energy consumption. Data for industry's share of energy consumption is available only from 1949 onward, and it shows that the industrial energy share has declined slowly over the last 60 years (data is from Energy Information Administration (EIA) Table 2.1). Between 1949 and 2000, the average annual change in the industrial energy share was  $-0.53\%$ . This is more than four times smaller than the  $2.2\%$  average annual change in total energy consumption during the same period. In order to maintain simplicity, I make the “not quite true” assumption that industrial energy share is approximately constant. Thus, its rate of change is approximately zero ( $\hat{f}_{IN} \approx 0$ ), meaning the growth rate of industrial output is approximately equal to the growth rate of fossil fuel output:

$$\hat{Q}_{IN} \approx \hat{Q}_M \quad (5.9)$$

**Table 5.1** Regressions for the output and productivity growth rates of the mining sector

$\hat{Q}_M =$	$-0.0588t + 118$
$\hat{\rho}_M =$	$-0.0179t + 37.6$

Since the only available biophysical productivity data comes from the mining sector, we assume that industrial labor productivity is proportional to the biophysical labor productivity of the mining sector ( $\hat{\rho}_M$ , fossil fuel output per worker), and that this proportionality is dictated by the constant  $\lambda$ . Under this assumption, the equation for the employment growth rate of industry becomes:

$$\hat{L}_{IN} = \hat{Q}_M - \lambda \hat{\rho}_M \quad (5.10)$$

Figure 5.6 plots the historical growth rates of fossil fuel output and mining sector biophysical labor productivity. Regressions for these two series are shown in Table 5.1. Note that the model uses *regression* data rather than raw historical data. The value of  $\lambda = 1.15$  reproduces nineteenth and early twentieth century industrial employment quite accurately (see Fig. 5.7, upper right). According to the model, this means that the labor productivity of industry grew 15 % faster than the mining sector productivity (although such an interpretation depends on the validity of all model assumptions). After 1940, however, the model diverges rapidly from historical data. There are two factors that are likely at play in this divergence. First, the assumption that industrial output is constrained by domestic fossil fuel production is no longer valid after 1960, as the US became a major energy importer from this point onward. Second, it seems likely that the parameter value  $\lambda = 1.15$  is no longer valid after 1940 (meaning the industrial labor productivity growth rate changed relative to the mining productivity growth rate). This means that we need to recalibrate our model.

To gain more insight into the post-WWII period, we reset our model to the initial year 1949 (when reliable annual energy consumption data becomes available). We now constrain the industrial output growth rate by the societal energy *consumption* growth rate ( $\hat{E}$ ), rather than the fossil fuel *production* growth rate. Furthermore, since we have access to accurate data, we use raw data for the historical energy growth rate, rather than a regression. Our function for industrial employment growth is now given by Eq. 5.11. This time  $\lambda = 0.68$  reproduces post-1949 industrial employment fairly accurately (Fig. 5.7, lower right). Assuming the validity of the model, this means that from 1949 onward, industrial productivity grew 68 % as fast as mining productivity. Notice that cyclical fluctuations are reproduced—a very desirable feature that is perhaps surprising, given the simplicity of our model. Since our energy data is the source of these fluctuations, it gives us an indication that we are on the right track.

$$\hat{L}_{IN} = \hat{E} - \lambda \hat{\rho}_M \quad (5.11)$$

We now move on to the service sector. Again, our basic equation is given by Eq. 5.12.

$$\hat{L}_s = \hat{Q}_s - \hat{\rho}_s \quad (5.12)$$

Hypotheses 2 and 3 (productivity increases require an increase in energy use, which requires an increase in facilitation activity) imply that an increase in the productivity

of the hypercycle requires an increase in facilitation activity. We tested this prediction by looking at the growth of facilitation activity within the hypercycle itself (Fig. 5.5). We now broaden this finding to assume that increases in hypercyclic productivity require a general rise in facilitation activity across all of society. Thus, we assume that the demand for services (i.e. service “output”) is proportional to the productivity of the hypercycle ( $\rho_{HYP}$ ), and that this proportionality is dictated by the constant  $\gamma$ :

$$\hat{Q}_S = \gamma \cdot \hat{\rho}_{HYP}. \quad (5.13)$$

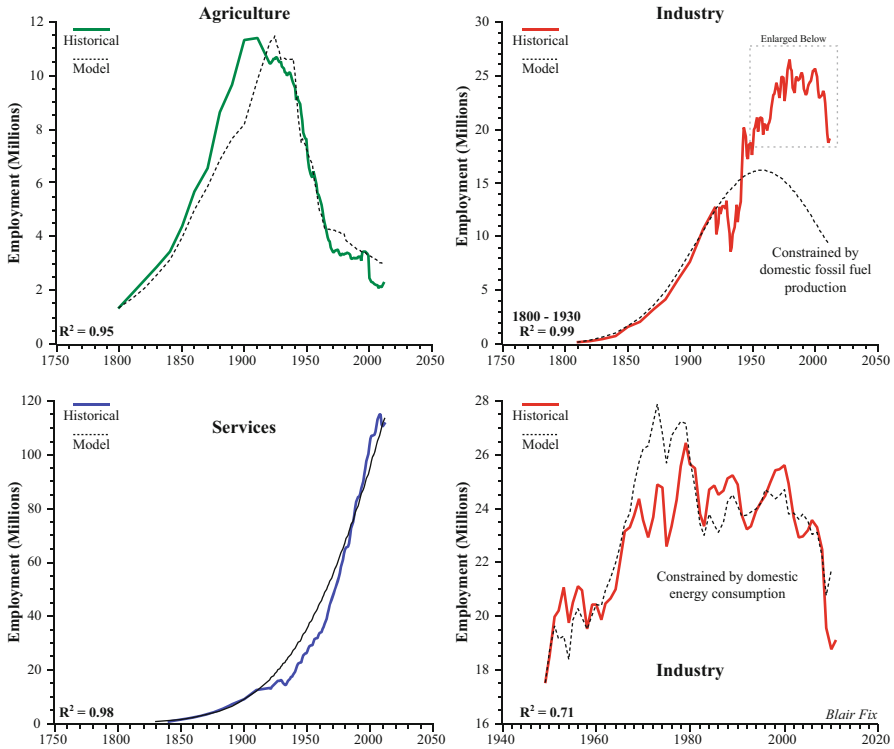
Furthermore, we assume that portions of the service sector (i.e. transportation) can increase their productivity, but that these productivity increases require a corresponding increase in facilitation activity. For instance, increases in transportation productivity may be the result of better infrastructure. However, a myriad of additional facilitation activity is required to design and maintain this infrastructure. Since all paid facilitation activity is contained within the service sector, productivity gains within this sector tend to be offset by a corresponding increase in a different facilitation activity. For simplicity, we assume that this process is “zero sum”, meaning productivity gains are exactly offset by increased facilitation activity. Under this assumption, service sector productivity is constant, thus its growth rate is zero ( $\hat{\rho}_S = 0$ ). Our equation for service sector employment growth becomes:

$$\hat{L}_S = \gamma \cdot \hat{\rho}_{HYP} \quad (5.14)$$

Since we want all exogenous inputs to the model to be biophysical, we proxy hypercycle productivity in terms of the biophysical productivity of the mining sector ( $\rho_M$ , fossil fuel output per worker). This yields Eq. 5.15. Note that we use the mining productivity *regression*, rather than raw historical data. The parameter value  $\gamma = 0.86$  allows us to fit historical service employment with an accuracy of  $R^2 = 0.98$  (see Fig. 5.7).

$$\hat{L}_S = \gamma \cdot \hat{\rho}_M \quad (5.15)$$

Having built a simple biophysical model that reproduces historical data, the important thing is to decide what the model *means*. In industry and agriculture, there are two basic components to the model. The first component is a constraint on output growth. In agriculture, the constraint is the food consuming ability of the total population. In industry, the constraint is total energy production/consumption. The second component is productivity growth. The dynamic of the model is produced by the interaction between the centripetal employment force of output growth and the centrifugal employment force of productivity growth. Both output growth and productivity growth are exogenous to the model, but use inputs of historical, biophysical data. In the service sector, the dynamic is simpler: the growth of service sector employment is proportional to the productivity growth of the hypercycle. The model shows that the simple dynamic between output constraints and productivity increases can account for most of the historical structural change in the US over the last two centuries.



**Fig. 5.7** Historical employment in agriculture, industry, and services and modeled results. (Sources: Employment data for 1800–1910 from HSUS Table Ba814-830. Non-agriculture employment data for 1920–2000 from HSUS Table Ba840-848. Agriculture employment data for 1920–1990 from HSUS Table Ba472; 1990–2000 from HSUS Table Ba482. Employment data for 2001–2010 from BEA Table 6.5D. US energy consumption from EIA Table 1.3. Total population from Maddison 2008. Wheat productivity is from HSUS Table Da1148. Industry includes results with and without imported energy)

The purpose for building this model is to understand how energy consumption is related to labor structure change. We can do this by returning to our initial hypotheses. According to Hypothesis 1, all sectors attempt to increase their labor productivity. According to Hypothesis 2, such increases require an increase in energy consumption per labor hour. Assuming all sectors (except services) increase their productivity, we can expect such productivity gains to lead to an overall increase in the per capita energy consumption of the entire society.

Thus, as an approximation, we can state that the productivity growth rate (for any given sector) is a function of the growth rate of societal energy use per capita ( $\hat{E}_{pc}$ ), and that this functional relation is governed by the parameter  $\mu$ . This parameter need not be constant (it could be a function of time) but it must remain *positive*, since we assume that productivity and societal energy consumption per capita are always

positively correlated. Thus for any nonservice sector  $x$ , we can write:

$$\hat{\rho}_x = \mu_x \hat{E}_x \quad (5.16)$$

We can use this relation to understand how our model relates to societal energy consumption. We begin in agriculture. Assuming that  $\hat{\rho}_{AG} = \mu_{AG} \hat{E}_{pc}$ , we can rewrite Eq. 5.4 by replacing the agricultural productivity growth rate with a function of the societal energy use per capita growth rate:

$$\hat{L}_{AG} = \hat{P}_T - \mu_{AG} \hat{E}_{pc} \quad (5.17)$$

Rearranging for energy use per capita gives:

$$\hat{E}_{pc} = \frac{\hat{P}_T - \hat{L}_{AG}}{\mu_{AG}} \quad (5.18)$$

Since  $\mu_{AG}$  is always positive, a consequence of Eq. 5.18 is that when the total population growth rate ( $\hat{P}_T$ ) exceeds the growth rate of employment in the agricultural sector ( $\hat{L}_{AG}$ ), the growth rate of societal energy use per capita will be positive. To state this result in a different way, a positive per capita energy growth rate (for all of society) should be accompanied by a decline in the size of agricultural employment relative to the size of the total population. Thus, we expect that societal energy use per capita should be negatively correlated with the relative size of the agricultural sector.

Next, we turn to the service sector. The growth rate of service employment is dictated by the growth rate of hypercyclic productivity. Again, we assume that the hypercyclic productivity growth rate can be written as a function of societal energy use per capita ( $\hat{\rho}_{HYP} = \mu_{HYP} \hat{E}_{pc}$ ). Substituting this function into Eq. 5.14 gives the following rather direct relation:

$$\hat{E}_{pc} = \frac{\hat{L}_S}{\mu_{HYP}} \quad (5.19)$$

Since  $\mu_{HYP}$  is always positive, increases in service employment should be positively correlated with increases in per capita energy consumption. Of course, this was the basis of hypothesis three, so our model is just reflecting our underlying hypothesis that increases in energy consumption require an increase in facilitation activity.

Lastly, we come to industry. Curiously, despite the fact that industry is the most energy-intensive sector, the relation between industrial employment and societal energy consumption is rather complex. Again we assume that the growth rate of industrial productivity is a function of societal energy use per capita ( $\hat{\rho}_{IN} = \mu_{IN} \hat{E}_{pc}$ ). Substituting this relation into Eq. 5.5 gives:

$$\mu_{IN} \hat{E}_{pc} = \hat{Q}_{IN} - \hat{L}_{IN} \quad (5.20)$$

However, in our model, we assumed that the industrial output growth rate was equal to the total (societal) energy consumption growth rate ( $\hat{Q}_{IN} = \hat{E}$ ). Therefore, we

substitute this into Eq. 5.20. Notice also that societal energy consumption per capita is defined as total energy consumption divided by total population ( $E_{pc} = E/P_T$ ). Therefore, the growth rate of societal energy use per capita is equal to the growth rate of energy minus the growth rate of population ( $\hat{E}_{pc} = \hat{E} - \hat{P}_T$ ). Substituting this into Eq. 5.20 gives:

$$\mu_{IN}(\hat{E} - \hat{P}_T) = \hat{E} - \hat{L}_{IN} \quad (5.21)$$

This can be rearranged into the following form:

$$\hat{E}(\mu_{IN} - 1) = \mu_{AG}\hat{P}_T - \hat{L}_{IN} \quad (5.22)$$

Interestingly, under the special condition that  $\mu_{IN} = 1$ , this equation collapses to  $\hat{L} = \hat{P}_T$ . What does this mean? Under the condition that the growth rate of industrial productivity is *exactly* equal to the growth rate of societal energy use per capita, industrial employment grows at the same pace as total population. Under this condition, any increase in society's per capita consumption of industrial goods is exactly offset by increases in industrial labor productivity. Thus, the industrial sector need only maintain its share of employment.

This also leads to the unexpected result that the relative size of industry can become *independent* of energy consumption. Under such a condition, there will be *no trend* between energy consumption and the employment share of industry. Since industry is by far the most energy-intensive sector, this result is counter-intuitive. Common sense leads to the conclusion that industrial employment share should be positively correlated with societal energy consumption.

The model developed above allows us to make two key predictions about the relation between per capita energy consumption and changes in labor structure:

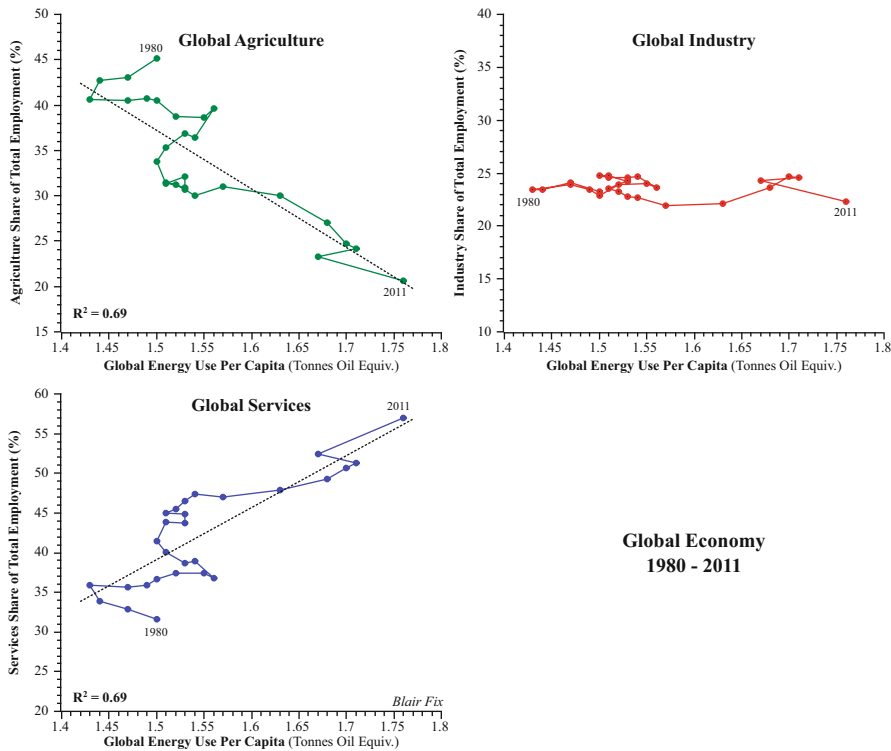
1. Agriculture employment share should be *negatively* correlated with energy use per capita.
2. Service employment share should be *positively* correlated with energy use per capita.

Furthermore, we can state that there is the *possibility* that the industrial employment share may have *no trend* with energy consumption.

## 5.5 Labor Structure and Energy: Empirical Evidence

We begin by noting that international trade makes it very difficult to conduct analysis of sectoral composition at the national level. One can never quite be sure how trade is affecting one's results. For instance work by Gierlinger and Krausmann (2012) and Giljum and Hubacek (2001) has shown that both the US and Europe now have major mass-balance deficits (they import more mass than they export). Thus, any conclusions drawn on the national level may rest on dubious grounds, since domestic



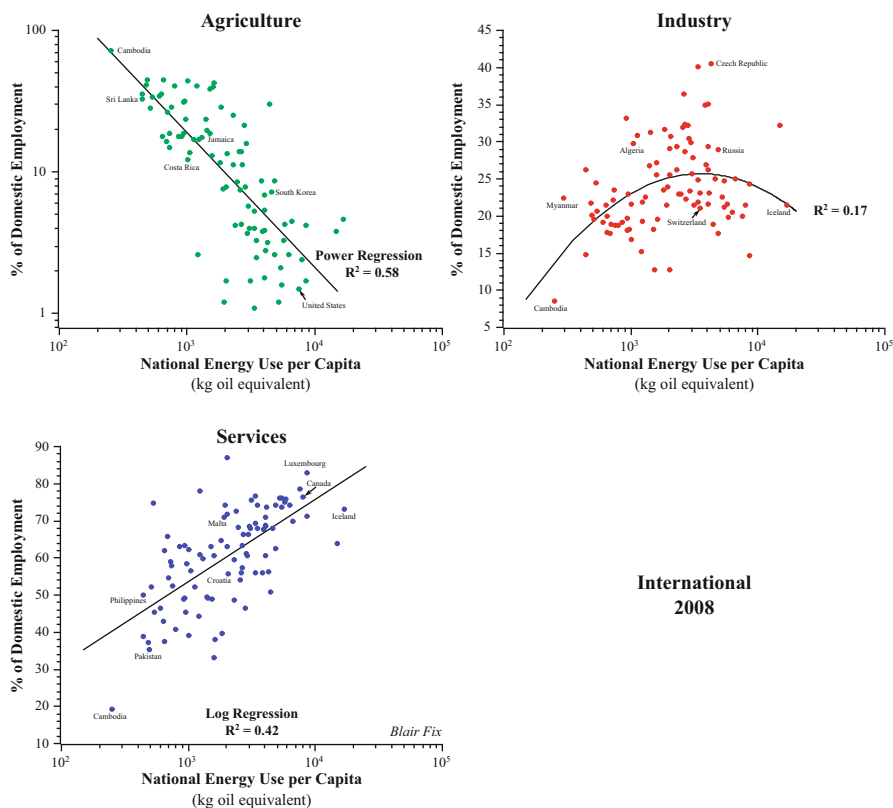


**Fig. 5.8** Global employment share of agriculture, industry and services vs. global energy consumption per capita. (Note: No causation is implied by the choice of y- and x-axes variables. Sources: All employment data from World Bank Development Indicators, data series SL.AGR.EMPL.ZS (Agriculture Share), SL.IND.EMPL.ZS (Industry Share), SL.SRV.EMPL.ZS (Service Share), and SP.POP.TOTL (National Population). Global sector shares are the average of all national shares, weighted by each nation’s share of world population. Global energy consumption from BP Statistical Review 2013)

sectors may be proportionately larger/smaller than would otherwise be possible in a closed system.

For this reason, we begin our empirical analysis at the scale of the global economy, the only truly closed economy on the planet. In a sense, results garnered at this level are the most incorruptible, since one can never say that a trade imbalance distorted one’s data. Figure 5.8 plots the sector composition of the entire global economy against global energy consumption per capita. Global sector composition is calculated by a weighted average of all nations, where the sector composition of each nation is weighted by its total population.

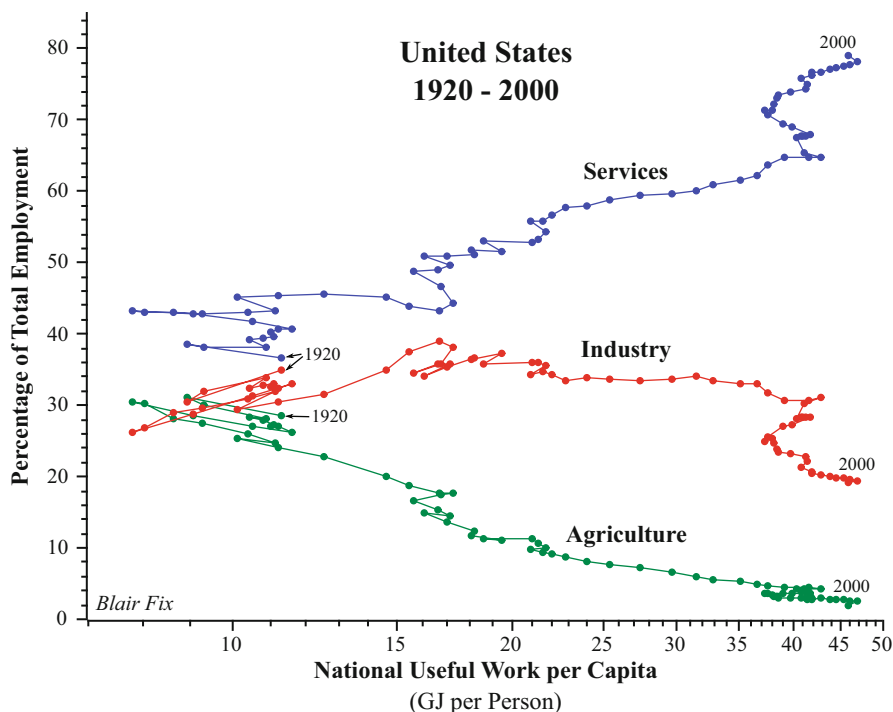
The results are consistent with the predictions made by our biophysical model: Increases in per capita energy consumption are associated with an increase in the service share of employment and a decrease in the agricultural share of employment. What is surprising, however, is that industrial employment share has virtually *no*



**Fig. 5.9** International comparison of employment shares of agriculture, industry and services vs. national energy consumption per capita. (Note: No causation is implied by the choice of y- and x-axes variables. Sources: All data from World Bank Development Indicators, data series SL.AGR.EMPL.ZS (Agriculture Share), SL.IND.EMPL.ZS (Industry Share), SL.SRV.EMPL.ZS (Service Share), and EG.USE.PCAP.KG.OE (Energy Use per capita))

*trend* with global energy consumption. In our model, this was predicted to occur under the stringent condition that the industrial productivity growth rate was *exactly* equal to the growth rate of societal energy use per capita. Thus, what was a possible, but *improbable* scenario in our model turns out to be a rather robust feature of the global economy.

We turn now to a cross-sectional analysis of the global economy in the year 2008. We plot the agricultural, industrial, and service employment share of each nation against national energy consumption per capita. By doing so, we are able to see how the per capita energy consumption of each nation relates to its sectoral composition (Fig. 5.9). Again, the results are consistent with our predictions: agricultural employment share is negatively correlated with energy consumption, while service employment share shows a positive correlation. Interestingly, the relation between industrial employment and energy use is extremely muddy. It does seem to resemble



**Fig. 5.10** US employment share of agriculture, industry and services vs. national useful work per capita. (Note: No causation is implied by the choice of y and x-axes variables. Sources: Nonagriculture employment data for 1920–2000 from HSUS Table Ba840-848. Agriculture employment data for 1920-1990 from HSUS Table Ba472; 1990–2000 from HSUS Table Ba482. Useful work data from Benjamin Warr REXS Database)

an inverted U-shape, but the correlation is very low. It is quite plausible that trade imbalances have influenced the shape of this curve.

Lastly, we move to a time-series comparison at the national level. Figure 5.10 plots US sectoral composition against national useful work consumption per capita. Again, the results are consistent with our predictions: agriculture employment is negatively correlated with useful work, while the service sector is positively correlated. As with the global level, we find that for much of the period in question, US industrial employment share had essentially no trend with per capita useful work consumption.

Two important conclusions can be drawn from this data. First, if we accept energy consumption as a metric for biophysical scale, then it follows that growth and sector composition are *fundamentally* related, contrary to assumptions made by neoclassical growth theory. Second, these results contradict the common assertion that a transition to a service economy will lead to the use of fewer resources (Jackson 1996; Stahel 1997). Since the service sector has a lower resource intensity than the industrial sector, the argument is that increasing the service composition of the economy should lead to the use of fewer resources. The evidence, however, suggests that the exact

*opposite* is true: Long-term increases in service sector size lead to an *increase* in energy consumption.

As is clear in our biophysical model, the problem with trying to dematerialize through service expansion is that the service sector is not functionally viable on its own. It is the dissipation and facilitation component of an *industrial* economy. Thus, a large service sector must be supported by a highly productive industrial sector, and such productivity inherently requires vast inputs of energy and resources. It is interesting to note that the *only* energy reduction pathway that is supported by empirical evidence is a return to an agrarian economy.

## 5.6 Conclusions: Labor Structure Matters

The implicit assumption built into the one-sector neoclassical growth model is that labor structure is unimportant to growth. However, our empirical findings, coupled with our biophysical model, lead to exactly the opposite conclusion: Labor structure change is *fundamentally* connected to changes in energy consumption (our metric for biophysical growth).

The basic insight of our model is that productivity gains and output constraints can be used to provide a simple explanation of labor structure change. Productivity gains in the hypercycle (the portion of the economy that produces goods) are *always* met by an increase in facilitation activity (i.e., services). In agriculture, output is constrained by population, which grows much more slowly than agricultural productivity. Thus, decreases in the agricultural employment share go hand in hand with increases in agricultural productivity. In industry, the matter is complicated, but we found that both the global economy and the USA showed essentially no trend between energy use per capita and industrial labor share. This is consistent with the (improbable) modeled scenario in which the industrial productivity growth rate is equal to the growth rate of energy use per capita.

The key strength of this model lies in its explicit linkage between energy consumption and productivity. Since productivity gains can be tied to increases in energy consumption per worker, we can state that increases in energy consumption are the key engine of structural change. However, note that causation need not be linear. While we can state that the growth of energy consumption *allows* for structural change, we can also state that structural change is *required* for increases in energy consumption to occur. As Giampietro et al. (2012) noted, this circularity (or “impredicativity”) is a key feature of complex systems. It means that establishing linear causation becomes difficult, if not impossible.

More than 70 years ago, Colin Clark wrote that “the most important concomitant of economic progress” is “the movement of working population from agriculture to manufacture, and from manufacture to commerce and services” (1940, p. 176). Contrary to the neoclassical vision in which changes in sectoral composition are a “relatively unimportant byproduct of growth” (Echevarria, 1997, p. 431), the

empirical and theoretical work conducted in this chapter tends to support Clark's view.

Perhaps the most compelling reason for understanding sectoral change is that current trends will likely *not* continue indefinitely. The evidence suggests that the demographic inversion, that is characteristic of highly industrialized nations (Fig. 5.1), is possible only through an immense mobilization of mostly fossil fuel energy. Any decline in the energy supply will render this configuration inoperative. Given this impending future, it is of great importance that we understand the relation between energy consumption and sectoral composition. While the simple model of sectoral change used in this chapter was good for fitting historical data, it is not suitable for modelling the future. To do this, a systems model is needed—one that takes into consideration the interaction between sectors. Such a model should have only one exogenous input—the scale of the total fossil fuel stock.

However, there is one prediction that can be made without a model: Any significant and long-lasting decline in per capita energy consumption will likely be associated with an increase in the agricultural share of employment and a decline in the service share of employment.

## References

- Baily MN, Gordon RJ, Solow RM (1981) Productivity and the services of capital and labor. *Brook Pap Econ Act* 1981(1):1–65
- Baran PA, Sweezy PM (1966) *Monopoly capital: an essay on the American economist and social order*. Monthly Review Press, New York
- Baumol WJ (1967) Macroeconomics of unbalanced growth: the anatomy of urban crisis. *Am Econ Rev* 57(3):415–426
- BLS (2009) BLS handbook of methods. <http://www.bls.gov/opub/hom/>
- Carter SB, Gartner SS, Haines MR, Olmstead AL, Sutch R, Wright G (2006) *Historical statistics of the United States: millennial edition*. Cambridge University Press, Cambridge
- Chinloy P (1980) Sources of quality change in labor input. *Am Econ Rev* 70(1):108–119
- Clark C (1940) *The conditions of economic progress*. Macmillan, London
- Cleveland C, Costanza R, Hall C, Kaufmann R (1984) Energy and the US economy: a biophysical perspective. *Science* 225(4665):890–897
- Denison EF (1962) *Sources of economic growth in the United States and the alternatives before us*. Committee for Economic Development, New York
- Echevarria C (1997) Changes in sectoral composition associated with economic growth. *Int Econ Rev* 38(2):431–452
- Eigen M (1971) Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58(10):465–523
- Engel E (1857) Die productions-und consumptions-verhältnisse des königreichs sachsen. *Bulletin de l'Institut International de la Statistique* 9
- Felipe J, Fisher FM (2003) Aggregation in production functions: what applied economists should know. *Metroeconomica* 54(2–3):208–262
- Fisher AGB (1935) *The clash of progress and security*. Macmillan, London
- Foellmi R, Zweimüller J (2008) Structural change, Engel's consumption cycles and Kaldor's facts of economic growth. *J Monetary Econ* 55(7):1317–1328
- Francois JF (1990) Producer services, scale, and the division of labor. *Oxf Econ Pap* 42:715–729

- Giampietro M, Mayumi K, Sorman A (2012) *The metabolic pattern of societies: where economists fall short*. Routledge, New York
- Giampietro M, Mayumi K, Sorman A (2013) *Energy analysis for a sustainable future: multi-scale integrated analysis of societal and ecosystem metabolism*. Routledge, New York
- Gierlinger S, Krausmann F (2012) The physical economy of the United States of America. *J Ind Ecol* 16(3):365–377
- Giljum S, Hubacek K (2001) International trade, material flows and land use: developing a physical trade balance for the European Union. Interim Report International Institute for applied systems analysis (01–059)
- Greenfield HI (1966) *Manpower and the growth of producer services*. Columbia University Press, New York
- Greenhalgh C, Gregory M (2001) Structural change and the emergence of the new service economy. *Oxf Bull Econ Stat* 63(s1):629–646
- Guilford MC, Hall C, O'Connor P, Cleveland C (2011) A new long term assessment of energy return on investment (EROI) for US oil and gas discovery and production. *Sustainability* 3(10):1866–1887
- Hall CAS, Cleveland CJ, Kaufmann RK (1986) *Energy and resource quality: the ecology of the economic process*. Wiley Interscience, New York
- Hodgson GM (2012) On the limits of rational choice theory. *Economic Thought: History, Philosophy, and Methodology* 1(1)
- Jackson T (1996) *Material concerns: pollution, profit, and quality of life*. Routledge, New York
- Leontief W (1971) Theoretical assumptions and nonobserved facts. *Am Econ Rev* 61(1):1–7
- Maddison A (2008) Statistics on world population, GDP and per capita GDP, 1–2008 AD. <http://www.ggdc.net/maddison/Maddison.htm>
- Ngai R, Pissarides CA (2004) Structural change in a multi-sector model of growth. Center for Economic Performance Discussion Paper No. 627
- North DC (1990) *Institutions, institutional change and economic performance*. Cambridge University Press, New York
- Perry GL, Denison EF, Solow RM (1971) Labor force structure, potential output, and productivity. *Brookings Pap Econ Act* (3):533–578
- Schwartzman D (1968) The contribution of education to the quality of labor 1929–1963. *Am Econ Rev* 58(3, Part 1):508–514
- Stahel WR (1997) The service economy: ‘wealth without resource consumption’? *Philos Trans Royal Soc Lond Ser A: Math, Phys Eng Sci* 355(1728):1309–1319
- Ulanowicz RE (1986) *Growth and development: ecosystems phenomenology*. Springer, New York
- Warr B (2009) REXS database. <https://sites.google.com/site/benjaminwarr/the-economic-growth-engine/rexs-database>. Accessed 9 April 2012
- Whelan K (2000) Balanced growth revisited: a two-sector model of economic growth. Federal Reserve Board, Division of Research and Statistics working paper
- Whelan K (2003) A two-sector approach to modeling US NIPA data. *J Money, Credit Bank* 35(4):627–656
- Zipf GK (1941) *National unity and disunity*. Principia Press, Bloomington

## Chapter 6

# Toward a Biophysical Growth Theory

*We've been ... trained to think of economics as being about money, and to some degree it is. But fundamentally it's about stuff. And if it's about stuff, why are we studying it as a social science? Why are we not, at least equally, studying it as a biophysical science?*

Charles Hall (quoted in Inman 2013)

Having conducted an extensive investigation into the empirical realities of growth, we are now in a situation to make some important decisions. I argue that it is time to abandon neoclassical growth theory. Below, I review the basic problems with neoclassical theory. I then argue that a biophysical growth theory is needed and I present a set of “stylized facts” that such a theory must explain. I conclude by reviewing some of the basic reasons why such a theory is a timely endeavour.

### 6.1 The Problem with Neoclassical Growth Theory

This book tested four implicit assumptions made by neoclassical growth theory:

1. Economic output can become decoupled from energy inputs.
2. Economic distribution is unrelated to growth.
3. Large institutions are not important for growth.
4. Labor force structure is not important for growth.

In all cases, the empirical evidence directly contradicted these assumptions. For those who think that a scientific theory should be based on empirically grounded facts, this critique alone provides compelling reasons to abandon neoclassical growth theory. But while the specific arguments in this book are new, critiques of neoclassical assumptions are not. Indeed, voluminous literature exists that questions the veracity of neoclassical assumptions (for instance: Robinson 1953, 1962; Georgescu-Roegen 1971; Leontief 1971; Mirowski 1991; Hall et al. 2001; Hall and Klitgaard 2012; Keen 2001 Nitzan and Bichler 2009, 2011). However convincing one may find this literature, it has not convinced many neoclassical theorists to abandon their theory.

There are many possible reasons for the doggedness with which neoclassical theorists adhere to their theory. When faced with criticism of their theory, perhaps the most pervasive response given by the applied economists is simply that the theory “works”. Indeed, the one sector production function is ubiquitous in applied macroeconomics, and neoclassical economists would not continue to use it if it did not give “good results”. However, these “good results” in no way vindicate neoclassical theory.

The first aspect of these “good results” is that aggregate production functions can be used to accurately fit historical gross domestic product (GDP) data. Unfortunately, a simple exponential function of time ( $Y_t \propto e^{at}$ ) can also be used to fit such data. For instance, a function with a constant growth rate of 3.5 % fits historical US GDP with an accuracy of  $R^2 = 0.98$  (between 1929 and 2011, data from the Bureau of Economic Analysis (BEA)). But it would be ludicrous to conclude that growth is an *inevitable* function of time. And yet, inside almost every aggregate neoclassical production function is a technological progress term that is nothing but an exponential function of time. While this exponential function is attributed to “technological progress”, such a term cannot be measured independently of neoclassical theory (and is usually assigned to the unexplained residual in conventional Cobb–Douglas equations). Thus, under the logic “use something if it works”, Occam’s razor would suggest that we use a pure exponential function of time to explain GDP: we need not bother with capital and labor inputs at all.

The second aspect of these “good results” is that aggregate production functions “reinforce the marginal productivity theory . . . of distribution” (Solow 1987, pp. 16–17; cited in Shaikh 2005, p. 447). Neoclassical theory predicts that the exponents in the Cobb–Douglas production function should be equal to the share of labor and capital in the national income. Thus, “good results” mean that regressions for the Cobb–Douglas exponents are very close to the values predicted by neoclassical theory. Yet as the work of Fisher (1971), Shaikh (1974), and McCombie (2001) demonstrate, these “good results” occur because the Cobb–Douglas form is an algebraic manipulation of the national accounts *identity*—the formula used by statistical agencies to *define* GDP. Thus, Shaikh writes that it is “impossible to distinguish the neoclassical argument from a mere tautology” (2005, p. 449). Therefore, the “good results” of neoclassical growth theory mean precisely nothing.

To conclude, the shortcomings of neoclassical growth theory can be summarized as follows:

1. *It does not explain the phenomenon for which it is designed to explain.* The majority of growth is attributed to the “Solow-residual”, which is an internalized error function. Neoclassical economists model the residual with an exponential function of time. However, if resorting to a function of time, Occam’s razor would suggest that we discard the remainder of the production function in favor of a pure function of time.



2. *There are fundamental problems associated with the measurement of the theory's basic variables (output and capital input).* The accepted method is to measure capital and output quantities by way of monetary value. However, such an approach requires making inherently subjective decisions, since the underlying unit (price) is not well-defined. Moreover, it appears that the current approach to measuring capital and output may be circularly dependent on neoclassical theory. Therefore, such metrics are inappropriate for testing neoclassic theory.
3. *Its implicit assumptions are directly contradicted by empirical evidence.* Rather than being “innocuously” untrue, the implicit assumptions made by neoclassical theory are “insidiously” untrue. The theory excludes from its scope some of the most fundamental aspects of growth. Thus, neoclassical growth theory maintains simplicity by courting irrelevance.
4. *The remaining empirical support for the theory is tautological.* The strong empirical results on which neoclassical growth theory purportedly rests neither elucidate the underlying technical form of the economy nor provide support for the marginal productivity theory of distribution. Instead, they are the result of a tautological relation between the production function form and an algebraic transformation of the national accounts identity.

Where, then, does this leave neoclassical growth theory? It seems fair to conclude that it is an elegant mathematical construct that has little to do with the real world.

## 6.2 A Biophysical Approach

Given the inadequacy of neoclassical theory, what is the best alternative? As should be obvious by now, I think that a biophysical approach to growth theory provides the most suitable way forward. Such a theory must begin by asking a very simple question, but one that is not often asked in economic theory: Why do we have growth *at all*? Indeed, growth is such an ephemeral phenomenon in the history of humanity that its very existence should be surprising. In my opinion, satisfying theories about the origins of growth do not come from economics, but from thermodynamics and the study of complex systems. In order to understand why growth exists, I propose that we need only two hypotheses:

1. All complex, nonequilibrium systems must be sustained by flows of energy and/or matter. Increases in these flows allow the system to expand.
2. An industrial economy is a nonequilibrium system that is energetically sustained primarily by exploitation of the finite stock of fossil fuels.

Growth, then, is possible whenever a new energy source is made available. Prior to industrialization, technological constraints prevented humans from exploiting fossil fuel energy. However, once sufficient technology existed, a feedback loop set in. Previously harvested resources and energy were transformed into technology that was powered by fossil fuels and which generated enough surplus to not only power the economy but also to exploit further fuels. Continuous iteration of this loop led to exponential growth.

What is the simplest way to model this feedback loop? Bardi and Lavacchi (2009) have shown that the famous Lotka–Volterra equations (which are usually used to model predator–prey dynamics) can be adapted to model the resource exploitation process:

$$\dot{R} = -k_1 T R \quad (6.1)$$

$$\dot{T} = k_2 T R - k_3 T \quad (6.2)$$

Here  $R$  represents a resource stock and  $T$  represents a stock of technological infrastructure. Equation 6.1 states that the rate at which the resource is harvested ( $\dot{R}$ ) depends upon the size of the technological stock, the size of the resource stock, and the efficiency of resource extraction ( $k_1$ ). This equation indicates that a greater technological stock can accelerate resource exploitation, but as the size of the resource stock dwindles (as  $R$  decreases), the pace of resource exploitation will slow.

Equation 6.2 states that harvested resources are transformed into technology. The rate of this transformation ( $\dot{T}$ ) is dictated by the rate of resource harvest ( $T R$ ) and the efficiency of the transformation process ( $k_2$ ). Additionally, technology (and its instruments) is subject to entropic decay ( $-k_3 T$ ) at a rate determined by  $k_3$ .

Biophysical growth, as I have defined it in this book, is the increase in the rate at which resources (specifically energy) flow through the economy. Thus, in Bardi and Lavacchi’s model, biophysical growth is represented by the rate of resource extraction ( $\dot{R}$ ). A robust feature of this model is that it produces bell-shaped resource extraction curves through time (like the one shown in Fig. 1.1). Thus, the essential insights of this model are: (1) growth can be modeled in terms of a feedback loop between technology and natural resource extraction; and (2) the ultimate growth limit is set by the size of the finite stock of resources.

This model gives some analytic rigor to the peak and decline scenario envisioned at the outset of the book. But while it indicates that a future energy consumption curve might be bell-shaped, it does not indicate how a future energy decline will affect society. It is also important to distinguish between external and internal constraints to growth. External (resource) constraints can describe the long run behavior of the economy, but internal (social) constraints dominate the short run. Historical crises have almost all been due to internal, social dynamics (think of the Great Depression). Even with the imposition of external, biophysical constraints, there is little reason to think that complex social dynamics will cease to be of importance in the future. Thus, an understanding of the future will require models, but also in-depth empirical study of the past.

What is needed is a biophysical research agenda—one that seeks to systematically understand the relation between energy consumption and all aspects of human society. Energy scholars such as Ayres and Warr (2009), Giampietro et al. (2012), Hall and Klitgaard (2012), and Smil (2010) have made significant contributions on this front, but much more work is needed.

## 6.3 Stylized Biophysical Facts

As I stated at the outset of the book, a good starting point for a new theory is to investigate the assumptions made by existing theory. If the results of this book tell us nothing else, it is that a good starting place for a biophysical growth theory is to begin with what neoclassical theory *ignores*. Neoclassical growth theory ignores the role of energy, yet the expansion of energy consumption is the single most important aspect of growth. Neoclassical growth theory ignores distribution, yet we found that distribution is fundamentally connected to growth. Neoclassical growth theory ignores large institutions, yet such institutions play a central role in growth. Neoclassical theory ignores changes in labor structure, yet changes in this structure are essential to growth.

A theory is always the product of the phenomena it seeks to explain. What does neoclassical growth theory seek to explain? Nearly 60 years ago, Nicholas Kaldor (1957) outlined six statements that came to be known as the “Kaldor facts” of economic growth. In many ways, the goal of neoclassical growth theory has been to explain these facts. Kaldor’s facts can be paraphrased as follows:

1. Output per worker grows at a roughly constant rate that does not decrease over time.
2. Capital per worker increases over time.
3. The capital/output ratio is roughly constant.
4. The rate of return to capital is roughly constant.
5. The share of capital and labor in net income are roughly constant.
6. Labor productivity growth rates vary considerably between societies.

Notice that five out of six of these facts are concerned with either something that remains “constant” (facts 1, 3, 4, 5) and/or something that “grows over time” (facts 1, 2). The logical offspring of these facts is a theory in which growth is constant and inevitable (i.e., neoclassical growth theory). Notice also the focus on *capital*. Neoclassical growth theory places capital at the center of its explanation of growth, but never bothers to explain *where capital comes from*. This neglect is likely the result of the neoclassical duality of capital (Nitzan and Bichler 2009). Note that when one applies compound interest to financial capital, the financial stock will grow exponentially. Neoclassical theory takes the logic of financial capital and applies it to the physical capital stock. Yet such a stock cannot be self-perpetuating—the laws of thermodynamics forbid it (Soddy 1926). Neoclassical theory fails to see that physical capital (i.e., a technological stock) is primarily a means for converting energy into useful work. Without an energy flow, physical capital cannot fulfill its purpose (think of a tractor without fuel).

By focusing on constant and inevitable growth driven by the accumulation of capital, neoclassical theory set itself on the wrong course from the very beginning. The focus of a growth theory should be on *energy*. Energy is the driving force that sustains all biophysical systems. While natural scientists have known this for years, theoretical inertia has prevented this simple fact from penetrating economic theory.

But as Hall et al. note, “economics is too important to leave to the economists” (2001, p. 663). It is time to begin constructing a growth theory that has a firm biophysical basis.

The goal of a biophysical growth theory should be to understand how changes in energy consumption affect society. In an attempt to focus the development of such a theory, I use the empirical evidence found in preceding chapters to propose the following seven stylized biophysical facts.

### Seven Stylized Biophysical Facts:

Trends accompanying increases in energy use per capita:

1. Large institutions (corporations and governments) increase their employment share (Sect. 4.3).
2. Agricultural employment decreases (Sect. 5.5).
3. Service employment increases (Sect. 5.5).

Trends accompanying increases in the energy use per capita *growth rate*:

1. The value of production increases relative to the price of energy (Sect. 2.4).
2. The share of profit in national income increases (Sect. 3.2).
3. Debt claims decrease relative to the value of production (Sect. 3.3).
4. Downward income redistribution is more likely to occur (Sect. 3.4).

A good starting point for a biophysical growth theory is to attempt to explain these seven stylized biophysical facts in a way that is both internally coherent and consistent with accepted scientific knowledge. I have offered my own thoughts and theories on the trends presented above, but the truth is that much remains unexplained. Acknowledging our ignorance, however, is the first step of any fruitful scientific inquiry.

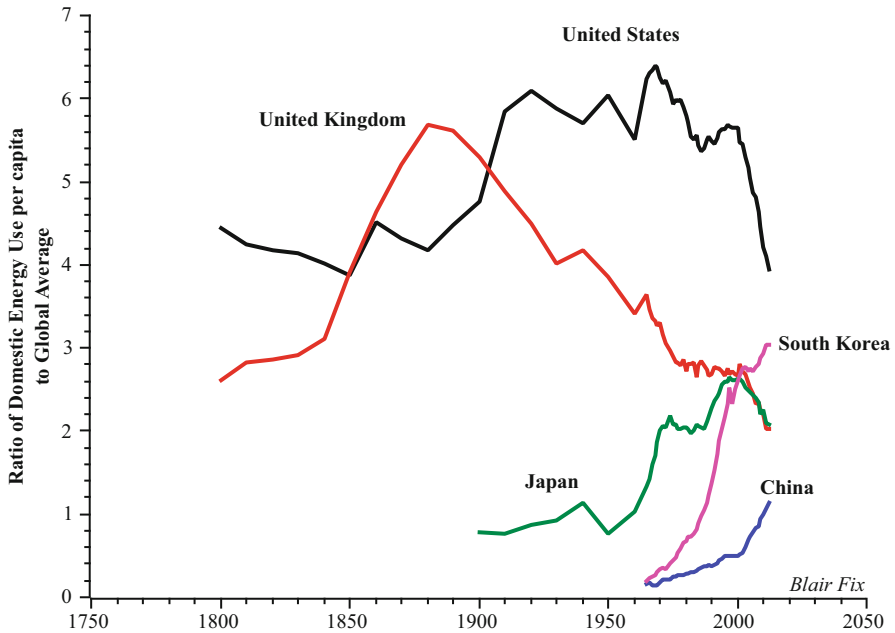
## 6.4 Conclusions

Supposing one accepts my assertion that neoclassical growth theory does not describe reality, and one accepts the hypothesis that energy and material flows (and the technology that enables them) form the basis for all growth, one could still object to the need for a biophysical growth theory on a number of grounds.

First, despite the fact that fossil fuels are finite on any meaningful human timescale, one might argue that their scale is so vast that the peak of production is a distant worry. In order to make meaningful predictions about such a peak, we need to know both the size and quality of the total fossil fuel stock. This is a difficult business that is currently the topic of a lively debate in energy literature. While recent empirical work suggests that the peak of conventional oil is occurring now<sup>1</sup> (Hall et al. 2014; Campbell 2009), estimates on recoverable coal vary widely Hook

---

<sup>1</sup> Technically speaking, a peak in production can be determined only in hindsight. Therefore, the most accurate statement we can make is that oil production is growing much, much more slowly than in the past.



**Fig. 6.1** National energy use per capita as a multiple of the global average. (Global energy consumption (1800–1960) from Vaclav Smil (2010). All non-US data after 1965 from BP Statistical Review of Energy 2013 (1965–present). US data from EIA Table E1 (1795–1945) and Table 1.3 (1949–2011). UK data from Paul Warde (2007). Population data from Maddison (2008) (1800–1960) and World Bank Development Indicators (1961–2010))

et al. (2010); Mohr and Evans (2009); Patzek and Croft (2010); Rutledge (2010). Therefore, it is not quite clear that a peak in total fossil fuel production is imminent (although, it does seem quite clear that it is occurring for higher quality fuels and for many geographical areas; see Hall and Klitgaard 2012).

Second, one might object that I have written off renewable energy. The future of renewable energy is difficult to predict. To date, though, renewable energy consumption (solar photovoltaic (PV) and wind) remains trivial—about 1 % of total primary energy consumption. However, it is conceivable that some game-changing, Promethean technology may arrive (like nuclear fusion, or a tremendous commitment to wind) that will completely replace fossil fuels. While I am not holding my breath, the possibility cannot be ruled out entirely.

Even if a forced energy decline is not imminent, I argue that there are still reasons why a biophysical growth theory is of immediate importance. Perhaps the foremost reason is that there is currently a fundamental shift occurring in the international pattern of energy consumption. Figure 6.1 shows historical per capita energy consumption of the UK, US, Japan, China, and Korea, expressed as a ratio of the global average.

In this chart we see the rise and relative decline of Britain during the nineteenth century, while we see the absolute dominance of US energy consumption rates during the twentieth century. However, as of the turn of the twenty-first century, US energy dominance is in unprecedented retreat. In fact, the major Western powers are now all experiencing a relative decline. On the other hand, Eastern nations, such as China and South Korea, are experiencing rapid increases in energy consumption. This trend has *immediate* repercussions, and a biophysical approach is absolutely necessary for making sense of it. However, quality biophysical research requires quality biophysical data. One simple policy recommendation that can be made by a biophysical growth theory is a reimplementation of a national program of collecting biophysical as well as monetary data on the economy on at least the scale and quality that the USA was undertaking routinely in the 1970s and 1980s.

On a more philosophical front, a biophysical growth theory is desirable because it starts from an energetics basis. As Chaisson (2002) notes, energy is the “universal currency” of the natural sciences. One of the great projects of science has been the unification of knowledge. A biophysical growth theory has the potential to offer a bridge between the natural and social sciences—something that is desirable in its own right.

Whether for practical or philosophical reasons, a biophysical growth theory is an important endeavor. The task now is to begin constructing it.

## References

- Ayres R, Warr B (2009) The economic growth engine: how energy and work drive material prosperity. Edward Elgar Pub, Northampton
- Bardi U, Lavacchi A (2009) A simple interpretation of hubbert’s model of resource exploitation. *Energies* 2(3):646–661
- Campbell CJ (2009) An atlas of oil and gas depletion. Jeremy Mills Pub., Huddersfield
- Chaisson EJ (2002) Cosmic evolution: the rise of complexity in nature. Harvard University Press, Cambridge
- Fisher FM (1971) Aggregate production functions and the explanation of wages: a simulation experiment. *Rev Econ Stat* 53(4):305–325
- Georgescu-Roegen N (1971) The entropy law and the economic process. Harvard University Press, Cambridge
- Giampietro M, Mayumi K, Sorman A (2012) The metabolic pattern of societies: where economists fall short. Routledge, New York
- Hall C, Klitgaard K (2012) Energy and the wealth of nations: understanding the biophysical economy. Springer, New York
- Hall C, Lindenberger D, Kümmel R, Kroeger T, Eichhorn W (2001) The need to reintegrate the natural sciences with economics. *BioScience* 51(8):663–673
- Hallock JL, Wu W, Hall CA, Jefferson M (2014) Forecasting the limits to the availability and diversity of global conventional oil supply: validation. *Energy* 64:130–153
- Hook M, Zittel W, Schindler J, Aleklett K (2010) Global coal production outlooks based on a logistic model. *Fuel* 89(11):3546–3558
- Hunt EK (2011) History of economic thought: a critical perspective. M.E. Sharpe, Armonk

- Inman M (2013) Will fossil fuels be able to maintain economic growth? a q&a with Charles Hall. *Sci Am* 308(4). <http://www.scientificamerican.com/article/eroi-charles-hall-will-fossil-fuels-maintain-economic-growth/>. Accessed 12 Aug 2014
- Kaldor N (1957) A model of economic growth. *Econ J* 67(268):591–624
- Keen S (2001) *Debunking economics: the naked emperor of the social sciences*. Zed Books, New York
- Leontief W (1971) Theoretical assumptions and nonobserved facts. *Am Econ Rev* 61(1):1–7
- Maddison A (2008) Statistics on world population, GDP and per capita GDP, 1-2008 AD. <http://www.ggdc.net/maddison/Maddison.htm>. Accessed 29 May 2013
- McCombie JSL (2001) What does the aggregate production function show? Further thoughts on Solow's "second thoughts on growth theory". *J Post Keynes Econ* 23(4):589–615
- Mirowski P (1991) *More heat than light: economics as social physics, physics as nature's economics*. Cambridge Univ Press, Cambridge
- Mohr S, Evans G (2009) Forecasting coal production until 2100. *Fuel* 88(11):2059–2067
- Nitzan J, Bichler S (2009) *Capital as power: a study of order and creorder*. Routledge, New York
- Patzek T, Croft G (2010) A global coal production forecast with multi-Hubbert cycle analysis. *Energy* 35(8):3109–3122
- Robinson J (1953) The production function and the theory of capital. *Rev Econ Stud* 21(2):81–106
- Robinson J (1962) *Economic philosophy*. Aldine Pub. Co., Chicago
- Rutledge D (2010) Estimating long-term world coal production with logit and probit transforms. *Int J Coal Geol* 85(1):23–33
- Shaikh A (1974) Laws of production and laws of algebra: the humbug production function. *Rev Econ Stat* 56(1):115–120
- Shaikh A (2005) Nonlinear dynamics and pseudo-production functions. *East Econ J* 31(3):447–466
- Smil V (2010) *Energy transitions: history, requirements, prospects*. Praeger Publishers, Santa Barbara
- Soddy F (1926) *Virtual wealth and debt: the solution of the economic paradox*. George Allen & Unwin, London
- Solow RM (1987) Second thoughts on growth theory. In: Steinherr A, Weiserbs D (eds) *Employment and growth: issues for the 1980s*, Springer, pp 13–28
- Warde P (2007) Energy consumption in England and Wales, 1560–2000. <http://www.histecon.magd.cam.ac.uk/history-sust/energyconsumption/>. Accessed 12 July 2012

# Appendix

## US Sectors: Methodology and Sources

### A.1 Data for Sectoral Corporate Concentration

#### A.1.1 *Corporate Employment*

Employment of top 100/200 corporations (ranked by number of employees) is from COMPUSTAT database (series DATA29). Primary and secondary sector is defined as SIC codes less than 3999, while the service sector is defined as SIC codes between 4000 and 8999.

#### A.1.2 *Total Sectoral Employment*

Data for primary and secondary sectors is from Bureau of Economic Analysis (BEA) Tables 6.5 B–D, full-time equivalent employees by industry. Employment data for private service sector from BLS online, series CES0800000001. See Fig. A.2 for the composition of these sectors.

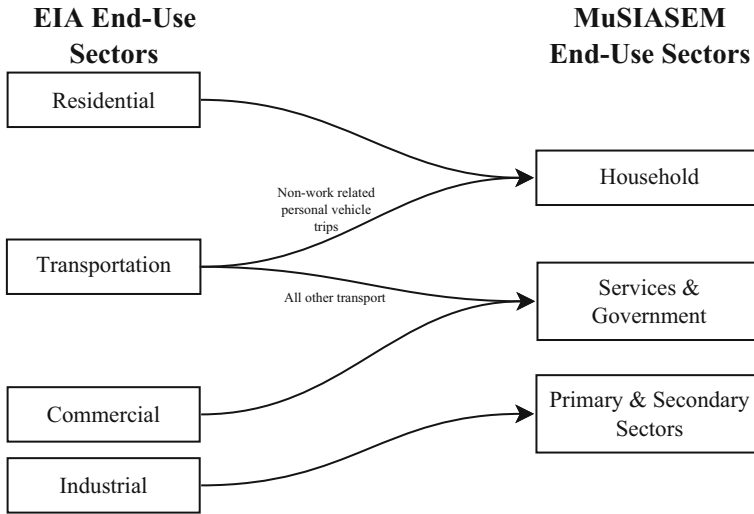
### A.2 Energy Consumption per Labor Hour

#### A.2.1 *Energy*

Sector energy consumption is from Energy Information Administration (EIA) Table 2.1a, energy consumption estimates by sector. However, following the methodology developed by Giampietro, Mayumi, and Sorman (in *The Metabolic Pattern of Societies: Where Economists Fall Short*) the EIA categories are reorganized into more functional end-use categories (Fig. A.1). Giampietro et al. call this the MuSIASEM approach (multiscale integrated analysis of societal and ecosystem metabolism).

This recategorization mainly involves the reallocation of transportation sector energy consumption to the household and service sector. Energy used for nonwork





**Fig. A.1 Recategorizing energy consumption**

related personal vehicle use is allocated to the household sector and is calculated using Eq. A.1 (where VMT stands for vehicle mile trips).

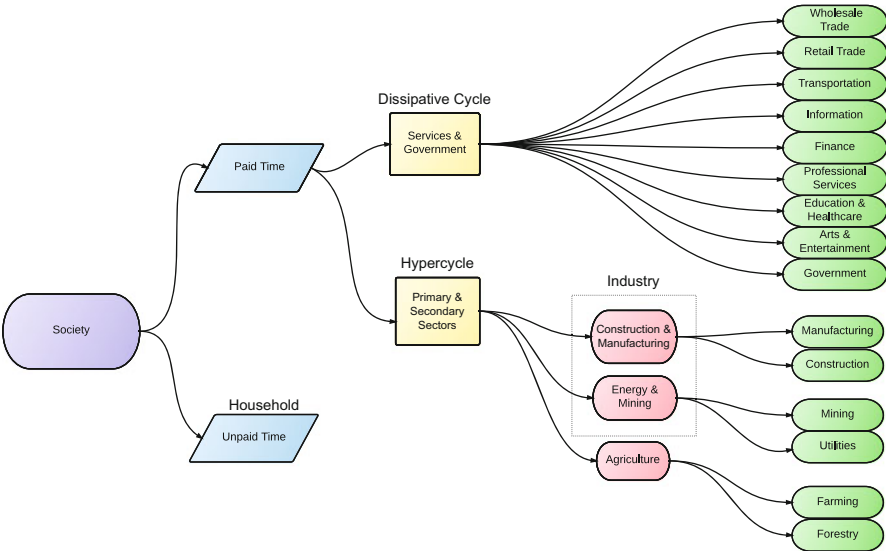
$$\text{Household transportation} = \frac{\text{Nonwork related VMT}}{\text{Total VMT}} \cdot (\text{Personal vehicle energy use}) \quad (\text{A.1})$$

All remaining transportation energy is allocated to the service sector. Data for vehicle mile trips is from the 2009 National Household Travel Survey, Table 6. Data between points is interpolated linearly while data prior to 1969 is extrapolated linearly from the 1969–1983 trend. Data for personal vehicle energy consumption for the period 1970–2010 is derived from numerous editions of the EIA *Annual Energy Outlook* using light-duty vehicle energy consumption. Data for the period 1949–1969 uses fuel consumption for passenger car and four-tire vehicles from the Department of Transportation, Table VM-201A.

Note: The category *four-tire vehicles* is introduced in 1966 at a nonzero value. Data prior to 1966 is estimated by exponential extrapolation.

## A.2.2 Labor Hours

Sectoral labor hours are from BEA Tables 6.8 A–D (Persons engaged in production). All values are indexed to Table 6.8 C and multiplied by  $40 \text{ h/week} \times 52.1775 \text{ weeks/year}$ . Note that the service sector is defined to include



**Fig. A.2 Mapping MuSIASEM sectors onto subsectors**

government. This is because energy statistics do not distinguish between private and public consumption. See Fig. A.2 for MuSIASEM sectors onto subsectors.