NBER WORKING PAPER SERIES

TOTAL FACTOR PRODUCTIVITY: A SHORT BIOGRAPHY

Charles R. Hulten

Working Paper 7471
http://www.nber.org/papers/w7471

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2000

The views expressed herein are those of the author and not necessarily those of the National Bureau of Economic Research.

© 2000 by Charles R. Hulten. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

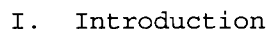
Total Factor Productivity: A Short Biography
Charles R. Hulten
NBER Working Paper No. 7471
January 2000
JEL No. O47, B22

ABSTRACT

Economists have long recognized that total factor productivity is an important factor in the process of economic growth. However, just how important it is has been a matter of ongoing controversy. Part of this controversy is about methods and assumptions. Total factor productivity growth is estimated as a residual, using index number techniques. It is thus a "measure of our ignorance," with ample scope for measurement error. Another source of controversy arises from sins of omission, rather than commission. A New Economy critique of productivity points to unmeasured gains in product quality, while an environmental critique points to the unmeasured costs of growth.

This essay is offered as an attempt to address these issues. Its first objective is to explain the origins of the growth accounting and productivity methods now under scrutiny. It is a biography of an idea, is intended to show what results can be expected from the productivity framework and what cannot. The ultimate objective is to demonstrate the considerable utility of the idea, as a counter-weight to the criticism, often erroneous, to which it has been subjected. Despite its flaws, the residual has provided a simple and internally consistent intellectual framework for organizing data on economic growth, and has provided the theory to guide a considerable body of economic measurement.

Charles R. Hulten
Department of Economics
University of Maryland
College Park, MD 20742
and NBER
hulten@econ.umd.edu



I. Introduction

Colonial Americans were very poor by today's standard of poverty. On the eve of the American Revolution, GDP per capita in the U.S. stood at approximately $765 in 1992 dollars.[1] Incomes rose dramatically over the next two centuries, propelled upward by the Industrial Revolution, and by 1997, GDP per capita had grown to $26,847. This growth was not always smooth (viz. Figure 1), but it has been persistent at an average annual growth rate of 1.7 percent. Moreover, the transformation wrought by the Industrial Revolution moved Americans off the farm to jobs in the manufacturing and (increasingly) in the service sectors of the economy.

The task of understanding this great transformation is one of the basic goals of economic research. The theorist has responded with a variety of models. Marxian and neoclassical theories of growth assign the greatest weight to productivity improvements driven by advances in the technology and organization of production. On the other hand, the New Growth Theory and

[1] Estimates of real GDP per capital and total factor productivity referred to in this section are pieced together from Gallman (1987), the Historical Statistics of the United States, Colonial Times to 1970, and the 1998 Economic Report of the President.

another branch of neoclassical economics -- the theory of capital and investment -- attach primary significance to the increase in investments in human capital, knowledge, and fixed capital.

The dichotomy between "technology" and capital formation carries over to empirical growth analysis. Generally speaking, the empirical growth economist has had two main tasks: first, to undertake the enormous job of constructing historical data on inputs and outputs; and, second, to measure the degree to which output growth is, in fact, due to technological factors ("productivity") and how much should be assigned to capital formation. This last undertaking is sometimes called "sources of growth analysis" and is the intellectual framework of the total factor productivity residual, which is the organizing concept of this survey.

A vast empirical literature has attempted to sort out the capital-technology dichotomy, but no clear consensus has emerged. Many of the early studies favored productivity as the main explanation of output growth (see Griliches (1996)), and this view continues in the "official" productivity statistics produced by the Bureau of Labor Statistics. However, Jorgenson and Griliches (1967) famously disagreed, and their alternative view finds support in subsequent work (e.g., Young (1995)) and in the New Growth Theory.

In recent years, attention has turned to another issue: the slowdown in productivity that started some time in the late 1960s or early 1970s. This issue has never been resolved satisfactorily, despite a significant research effort. This, in turn, has been supplanted by yet another mystery: why hasn't the widely touted information revolution reversed the productivity slowdown? Robert Solow (1987) puts the proposition succinctly: "We can see the computer age everywhere but in the productivity statistics." However,

this "Solow Paradox" is not limited to computers. One might as well say that 'We see the new technology everywhere but in the productivity statistics.' According to Nordhaus (1997), official price and output data "miss the most important technological revolutions in economic history." Moreover, the Advisory Commission to Study the Consumer Price Index (1996) assigns an upward bias of 0.6 percentage points per year in the CPI as a result of missed quality improvement, with a corresponding understatement of quantity.

In this "New Economy" critique of productivity statistics, the growth path evident in Figure 1, impressive as it may seem, seriously understates the true gains in output per person occurring over the last two centuries. However, there is another New Economy paradox that has been largely overlooked: if the missed quality change is of the magnitude suggested above, the quality of the goods in past centuries -- and the implied standard of living -- must have been much lower than implied by official (and allegedly quality-biased) statistics (Hulten (1997)). Indeed, taken to its logical extreme, the correction of Figure 1 for quality bias would result in a quality-adjusted average income in 1774 which is dubiously small.[2]

A second line of attack on the New Economy view comes from environmentalists, who argue that GDP growth overstates the true improvement in economic welfare, because it fails to measure the depletion of natural resources and the negative spillover externalities associated with rapid GDP growth (Gore (1992), Robert Repeto et. al. (1996)). This attack has been broadened to include what are asserted to be the unintended consequences of

[2] If all prices (not just the CPI prices) grew at a rate that was actually lower by 0.6 percent than official price statistics, the corresponding quantity statistics would have an offsetting downward bias. If this bias occurred all the way back to 1774, real GDP per capita would have been $202 in that year, not $765.

the Industrial Revolution: poverty, urban decay, crime, loss of core values, etc. This view is represented by a statement which appeared on the cover of Atlantic Monthly, "The gross domestic product (GDP) is such a crazy mismeasure of the economy that it portrays disaster as gain."

In other words, conventional estimates of productivity growth are either much too large or much too small, depending on one's view of the matter. The truth undoubtedly lies somewhere between the two extremes, but where? This essay is offered as an attempt to answer this question. Its first objective is to explain the origins of the growth accounting and productivity methods now under attack. This explanation, a biography of an idea, is intended to show what results can be expected from the productivity framework and what cannot. The ultimate objective is to demonstrate the considerable utility of the idea, as a counter-weight to the criticism, often erroneous and sometimes harsh, to which it has been subjected. The first part of the essay is a critical bibliography of the research works that have defined the field. A second part includes a somewhat personal tour of recent developments in the field and tentative answers to some of the unresolved issues.

II. The "Residual": A Critical Bibliography to the Mid-1980s

A. National Accounting Origins

Output per unit input, or total factor productivity, is not a deeply theoretical concept. It is, in fact, an implicit part of the circular income flow model familiar to students of introductory economic theory. In that model, the product market determines the price, p_t, and quantity, Q_t, of goods and services sold to consumers. The total value of these goods is p_tQ_t

dollars, which is equally the expenditure of consumers and the revenue of producers. The factor markets determine the volume of the inputs (labor, L_t, and capital, K_t), as well as the corresponding prices, w_t and r_t. The payment to these inputs, $w_tL_t+r_tK_t$, is a cost to the producer and the gross income of consumers. The two markets are connected by the equality of revenue and cost, on the producer side, and gross income and expenditure on the consumer side, leading to the fundamental GDP accounting identity:

$$p_tQ_t = w_tL_t + r_tK_t \quad . \tag{1}$$

This is, in effect, the budget constraint imposed on an economy with limited resources of capital, labor, and technology.

However, GDP in current prices is clearly an unsatisfactory metric of economic progress. Economic well-being is based on quantity of goods and services consumed, and not the amount spent on these goods. Since the volume of market activity as measured by the identity (1) can change merely because prices have risen or fallen, it can be a misleading indicator of economic progress. What is needed is a parallel accounting identity which records the volume of economic activity that holds the price level constant, i.e., a revision of the identity (1) using the prices of some base-line year for valuing current output and input.

The construction of a parallel constant-price account is a deceptively simple undertaking. If constant dollar value of output is equal to the constant dollar value of input in any one year, the equality cannot hold in the following year if an improvement in productivity allows more output to be

obtained from a given quantity of inputs.[3] In order to bring the two sides of the constant dollar account into balance, a scaling factor, S_t, is needed. The correct form of the constant-price identity is thus:

$$p_0 Q_t = S_t [w_0 L_t + r_0 K_t] \quad . \tag{2}$$

The scaling factor has a value of one in the base year 0, but varies over time as the productivity of capital and labor changes. Indeed, if both sides of the identity (2) are divided by $w_0 L_t + r_0 K_t$, it is apparent that the scaling factor S_t is the ratio of output per unit of total factor input.

Growth accounting is largely a matter of measuring the variable S_t and using the result to separate the growth of real output into an input component and a productivity component. Griliches (1996) credits the first mention of the "output per unit input" index to Copeland in 1937, followed by Copeland and Martin in 1938. The first empirical implementation of the output-per-unit input is attributed to Stigler (1947).

Griliches also observes that Milton Friedman uncovered one of the chronic measurement problems of productivity analysis -- the index number problem -- in his comment on the Copeland-Martin research. The problem arises because, with some rearrangement, (2) can be shown to be a version of the fixed-weight Laspeyres index:

---

[3] The basic problem is illustrated by the following situation. Suppose that output doubles from one year to the next while labor and capital are unchanged. If the accounting is done in the constant prices of the first year, the left-hand side of the constant price identity doubles while the right-hand side remains unchanged, violating the adding-up condition.

(3) $$\frac{S_t}{S_0} = \frac{\dfrac{Q_t}{Q_0}}{\dfrac{w_0L_t+r_0K_t}{w_0L_0+r_0K_0}} .$$

This is a widely used index formula (e.g., the Consumer Price Index) and was used in early productivity literature (e.g., Abramovitz (1956)). However, the substitution bias of the Laspeyres index is also well known (and recently pointed out by the Advisory Commission (1996) in its analysis of the CPI). Substitution bias arises when relative prices change and agents (producers or consumers, depending on the context) substitute the relatively cheaper item for the more expensive. The problem can sometimes be reduced by the use of chained (i.e., frequently reweighted) Laspeyres indexes, and both Kendrick (1961) and Denison (1962) endorse the use of chain-indexing procedures, although they primarily use fixed-weight procedures.

A more subtle problem arises in the interpretation of the ratio S_t. The basic accounting identities (1) and (2) can be read either from the standpoint of the consumer or from that of the producer. Virtually all productivity studies have, however, opted for the producer-side interpretation, as witnessed by terms like "output per unit input" and "total factor productivity." Moreover, discussions of the meaning S_t have typically invoked the rationale of the production function (see, for example, the long discussion in Kendrick (1961)). However, the consumer-welfare side has lurked in the background. The early literature tended to regard S_t as an indicator of the welfare benefits of innovation, with the consequence that "real" national income or real net national product was preferred to output measured gross of real depreciation when calculating the numerator of the total factor

productivity ratio.[4] This preference was based on the argument that an increase in gross output might be achieved by accelerating the utilization (and thus deterioration and retirement) of capital, thereby increasing total factor productivity without conveying a long-run benefit to society. This argument had the effect of commingling consumer welfare considerations with supply-side productivity considerations. This introduced a fundamental ambiguity about the nature of the total factor productivity index that has persisted to this very day in a variety of transmuted forms.

B. The Production Function Approach and the Solow Solution

Solow (1957) was not the first to tie the aggregate production function to productivity. This link goes back at least as far as Tinbergen (1942). However, Solow's seminal contribution lay in the simple, yet elegant, theoretical link that he developed between the production function and the index number approach. Where earlier index number studies had interpreted their results _in light_ of a production function, Solow started with the production function and deduced the consequences for (and restrictions on) the

---

[4] The concept of "depreciation" has been a source of confusion in the productivity and national income accounting literatures, and elsewhere (Hulten (1990), Triplett (1996)). Depreciation is a price concept, referring to the loss of capital _value_ because of wear, tear, obsolescence, and approaching retirement. The loss of productive capacity as a piece of capital ages is not, strictly speaking, depreciation. The capital stock loses capacity through in-place deterioration and retirement.

We will adopt the following terminology in this paper: the net value of output is the difference between the gross value and depreciation; real net output is the difference between constant-price (real) gross output and a constant-price measure of depreciation; net capital stock is the difference between the gross stock and deterioration.

productivity index. Specifically, he began with an aggregate production function with a Hicks' neutral shift parameter and constant returns to scale:

$$(4) \qquad Q_t = A_t F(K_t, L_t) \quad .$$

In this formulation, the Hicksian A_t measures the shift in the production function at given levels of labor and capital. It almost always identified with "technical change," but this is not generally an appropriate interpretation.[5]

Once the production function is written this way, it is clear that the Hicksian A_t and the ratio of output per unit of input S_t of the preceding section are related. The terms of the production function can be rearranged to express relative Hicksian efficiency, A_t/A_0, as a ratio with Q_t/Q_0 in the numerator and the factor-accumulation portion of the production function, $F(K_t,L_t)/F(K_0,L_0)$, in the denominator. The indexes A_t and S_t are identical in special cases, but A_t is the more general indicator of output per unit input (total factor productivity). In the vocabulary of index number theory, the Laspeyres S_t is generally subject to substitution bias.

Solow then addressed the key question of measuring A_t using a non-parametric index number approach (i.e., an approach that does not impose a

---

[5] The difference between the Hicksian shift parameter, A_t, and the rate of technical change arises for many reasons. The most important is that the shift parameter captures only costless improvements in the way an economy's resources of labor and capital are transformed into real GDP (the proverbial "Manna from Heaven"). Technical change that results from R&D spending will not be captured by A_t unless R&D is excluded from L_t and K_t (which it generally is not). A second general reason is that changes in the institutional organization of production will also shift the function, as will systematic changes in worker effort. We will emphasize these and other factors at various points throughout this paper.

specific form on the production function). The solution was based on the total (logarithmic) differential of the production function:

$$\frac{\dot{Q}_t}{Q_t} = \frac{\partial Q}{\partial K}\frac{K_t}{Q_t}\frac{\dot{K}_t}{K_t} + \frac{\partial Q}{\partial L}\frac{L_t}{Q_t}\frac{\dot{L}_t}{L_t} + \frac{\dot{A}_t}{A_t} \quad . \tag{5}$$

This expression indicates that the growth of real output on the left-hand side can be factored into the growth rates of capital and labor, both weighted by their output elasticities, and the growth rate of the Hicksian efficiency index. The former represent movements along the production function while the latter is the shift in the function.

The output elasticities in (5) are not directly observable, but if each input is paid the value of its marginal product, i.e., if

$$\frac{\partial Q}{\partial K} = \frac{r_t}{p_t}, \quad \frac{\partial Q}{\partial L} = \frac{w_t}{p_t} \quad , \tag{6}$$

then relative prices can be substituted for the corresponding marginal products. This, in turn, converts the unobserved output elasticities into observable income shares, s^K and s^L. The total differential (5) then becomes

$$\Re_t = \frac{\dot{Q}_t}{Q_t} - s^K{}_t \frac{\dot{K}_t}{K_t} - s^L{}_t \frac{\dot{L}_t}{L_t} = \frac{\dot{A}_t}{A_t} . \tag{7}$$

$\Re_t$ is the Solow residual: the residual growth rate of output not explained by the growth in inputs. It is a true index number, in the sense that it can be computed directly from prices and quantities. The key result of Solow's analysis is that $\Re_t$ is, in theory, equal to the growth rate of the Hicksian

efficiency parameter.

This is the theory. In practice, $\Re_t$ is a 'measure of our ignorance,' as Abramovitz (1956) put it, precisely because $\Re_t$ is a residual. This ignorance covers many components, some wanted (like the effects of technical and organizational innovation), others unwanted (measurement error, omitted variables, aggregation bias, model misspecification).

C. A Brief Digression on Sources of Bias

The unwanted parts of the residual might cancel if they were randomly distributed errors, leaving the systematic part of the residual unbiased. However, New Economy and environmentalist complaints arise precisely because the errors are thought to be systematic. These issues are on the agenda for the second half of this paper. Three other general criticisms will, however, be addressed here, in part because they involve challenges to the basic assumptions of the Solow model, and in part because they inform the evolution of the residual described in the next few sections.

First, there is the view that the Solow model is inextricably linked to the assumption of constant returns to scale. This view presumably originated from the close link between the GDP accounting identity (1) and the production function. If the production function happens to exhibit constant returns to scale <u>and</u> the inputs are paid the value of their marginal products as in (6), the value of output equals the sum of the input values. This "product exhaustion" follows from Euler's Theorem, and it implies that the value shares, s^K and s^L, sum to one. However, there is nothing in the sequence of steps (4) to (7), leading from the production function to the residual, that

requires constant returns (see Hulten (1973)). Constant returns is actually needed for another purpose: to estimate the return to capital as a residual, as per Jorgenson and griliches (1967). If an independent measure of the return to capital is used in constructing the share-weights, the residual can be derived without the assumption of constant returns.

A second general complaint against the residual is that it is married to the assumption of marginal cost pricing (i.e., to the marginal productivity conditions (6)). When imperfect competition leads to a price greater than marginal cost, Hall (1988) shows that the residual yields a biased estimate of the Hicksian shift parameter, A_t. There is, unfortunately, no way around this problem within the index number approach proposed by Solow. The index number approach is by nature non-parametric, meaning that it produces estimates of A_t directly from prices and quantities. The essence of the Solow method is to use prices to estimate the slopes of the production function at the observed input-output configurations, without having to estimate the shape of the function at all other points (i.e., without the need to estimate all the parameters of the technology). The residual is thus a parsimonious method for getting at the shift in the production function, but the price of parsimony is the need to use prices as surrogates for marginal products.

A third issue concerns the implied nature of technical change. In general, the Hicksian formulation of the production function (7) is valid if innovation improves the marginal productivity of all inputs equally. In this case, the production function shifts by the same proportion at all combinations of labor and capital. This is clearly a strong assumption which may well lead to biases if violated. A more general formulation allows (costless) improvements in technology to augment the marginal productivity of

each input separately:

$$(4') \qquad Q_t = F(a_t K_t, b_t L_t) \quad .$$

This is the "factor-augmentation" formulation of technology. It replaces the Hicksian A_t with two augmentation parameters, a_t and b_t. If all the other assumptions of the Solow derivation are retained, a little algebra shows that the residual can be expressed as

$$(7') \qquad \Re_t = s^K{}_t \frac{\dot{a}_t}{a_t} + s^L{}_t \frac{\dot{b}_t}{b_t} \quad .$$

The residual is now the share-weighted average of the rates of factor augmentation, but it still measures changes in total factor productivity. Indeed, when the rates of factor augmentation are equal, and the sum of the shares is constant, we effectively return to the previous Hicksian case.

Problems may arise if the rates of factor augmentation are not equal. In this situation, termed "Hicks-biased technical change," it is evident that productivity growth depends on the input shares as well as the parameters of innovation. A change in the income shares can cause output per unit input (total factor productivity) to increase, even if the underlying rate of technical change remains unchanged. This reinforces the basic point that productivity growth is not the same thing as technical change.

Some observers have concluded that the bias in technical change translates into a measurement bias in the residual (e.g., Rodrik (1998)). This is only true if one insists on identifying total factor productivity with technical change. However, the productivity residual does not "get off" free

and clear: factor-biased technical change may not lead to measurement, but it does generally lead to the problem of "path dependence" discussed in the following section.

D. The Potential Function Theorem

Solow's derivation of the residual deduces the appropriate index number formulation from the production function and, as a by-product, shows that it is not the Laspeyres form. But what type of index number is it? It was soon noted that (7) is the growth rate of a Divisia index (e.g., Richter (1966)), a continuous-time index related to the discrete-time chain index mentioned above. This linkage is important because it allows Solow's continuous formulation to be implemented using discrete-time data, while preserving the theoretical interpretation of the residual as the continuous shift in an aggregate production function.

However, this practical linkage has one potential flaw. Solow showed that the production function (4) and the marginal productivity conditions (6) lead to the growth rate form (7). He did not show that a researcher who starts with (7) will necessarily get back to the shift term A_t in the production function. Without such a proof, it is possible that the calculation (7) could lead elsewhere besides A_t, thus robbing the index of its conventional interpretation.

This issue was addressed in my 1973 paper, where it is shown that the Solow conditions are both necessary and sufficient. The expression (7) yields a unique index only if there is a production function (more generally, a "potential function") whose partial derivatives are equal to the prices used

to compute the index. The production function (cum potential function) is the "integrating factor" needed to guarantee a solution to (7), which is in fact a differential equation. If there is no production function, or if it is non-homothetic, the differential equation (7) cannot be (line) integrated to a unique solution. This problem is called "path dependence."[6]

The Potential Function Theorem imposes a good deal of economic structure on the problem in order to avoid path dependence. Unfortunately, these conditions are easily met. First, aggregation theory demonstrates that the necessary production function exists only under very restrictive assumptions (Fisher (1965)), essentially requiring all the micro production units in the economy (plants, firms, industries) to have production functions that are identical up to some constant multiplier (see also Diewert (1980)). If the aggregation conditions fail, a discrete-time version of the Divisia index might still be cobbled together, but the resulting numbers would have no unique link to the efficiency index A_t. Indeed, the theoretical meaning of A_t itself is ambiguous if the aggregation conditions fail.[7]

---

[6] The problem of path dependence is illustrated by the following example. Suppose that there is a solution to the Divisia line integral, but only for a particular path of output and input Γ_1 between points A and B. If a different path between these two points, Γ_2, gives a different value, then path dependence arises. If Divisia index starts with a value of 100 at the point A and the economy subsequently moves from A to B along Γ_1, and then back to A along Γ_1, the Divisia index will not return to 100 at A. Since the path can cycle between A and B along these paths, the index can, in principle, have a purely arbitrary value.

[7] Path dependence also rises if the aggregate production function exists but fails to satisfy any of the basic assumptions: marginal productivity pricing, constant returns to scale, and Hicksian technical change. This statement must, however, be qualified by the remarks of the preceding section. If an independent estimate of the return of capital is used when constructing the share weight of capital, s^K, then the Divisia productivity index is path independent even under non-constant returns to scale (Hulten (1973)). Moreover, if costless

When the Divisia index is path independent, Solow's procedures yield an estimate of the productivity residual which is uniquely associated with the shift in the production function. This result carries the important implication that the residual must be given a capacity interpretation, in this case, rather than a welfare interpretation. Or, more accurately, any welfare interpretation must be ancillary to this main interpretation. If, for example, the demand side of the circular flow diagram is invoked, and the potential function is taken to be the utility function of the consumer, the results must yield the same estimate of A_t.

The Potential Function Theorem also sheds light on the debate over net versus gross measures of output and capital. The theorem requires the units of output or input selected for use in (7) to be consistent with the form of the production function used as the integrating factor. To choose net output for computing the Solow residual, for example, is to assert that the production process generates net output from capital and labor, and that factor prices are equal to the net value of marginal product rather than the gross value of standard theory. This is an unusual view of real-world production processes, since workers and machines actually make gross units of output and the units of output emerging from the factory door are not adjusted for depreciation. Nor do we observe a price quoted for net output. Thus, the Potential Function Theorem comes down rather decisively on the side of gross output. Similar reasoning leads to the use of a net-of-deterioration concept of capital.

---

technical change is Harrod neutral, line integration of the residual $\Re$ is subject to path dependence, but integration of the ratio $\Re/s^L$ is not, and leads to a path independent index of the labor augmentation parameter, b_t, in (7'). The Divisia residual is more versatile than commonly believed.

E. Jorgenson-Griliches and Denison

The 1967 paper by Jorgenson and Griliches is a major milestone in the evolution of productivity theory. It advanced the hypothesis that careful measurement of the relevant variables should cause the Solow measure of total factor productivity to disappear. This is an intellectually appealing idea, given that the total factor productivity index is a residual "measure of our ignorance." Careful measurement and correct model specification should rid the residual of unwanted components and explain the wanted ones.

Jorgenson and Griliches then proceeded to introduce a number of measurement innovations into the Solow framework, based on a strict application of the neoclassical theory of production. When the renovations were complete, they found that the residual had all but disappeared. This result stood in stark contrast to prevailing results, which found that the residual did make a sizeable contribution to economic growth. However, this attack on (indeed, inversion of) the conventional wisdom was answered by Denison, whose own results were consistent with the conventional wisdom.

Denison (1972) compared his procedures with those of Jorgenson and Griliches and found that part of the divergence was caused by a difference in the time periods covered by the two studies, and that another part was due to a capacity utilization adjustment based on electricity use. The latter indicated a secular increase between equivalent years in the business cycle, and when this was removed, and the two studies put in the same time frame, Denison found that the Jorgenson-Griliches residual was far from zero.

The debate between Denison (1972) and Jorgenson and Griliches (1967, 1972) focused attention on the bottom line of empirical growth analysis: how

much output growth can be explained by total factor productivity (Manna from Heaven) and how much had to be paid for capital formation. However, in doing so, it obscured the true contribution of the Jorgenson-Griliches study, which was to cement the one-to-one link between production theory and growth accounting. For Solow, the aggregate production function was a parable for the measurement of total factor productivity, for Jorgenson and Griliches it was the blueprint. Implementing this blueprint led to a number of important innovations in the Solow residual -- a sort of productivity improvement in the total factor productivity model itself.

One of the principal innovations was to incorporate the neoclassical investment theory developed in Jorgenson (1963) into productivity analysis. The first step was to recognize that the value of output in the accounting identity (1) is the sum of two components: the value of consumption goods produced, $p^C C$, and the value of investment goods produced, $p^I I$ (hence, $pQ = p^C C + p^I I = wL + rK$). The price of the investment good was then assumed to be equal to the present value of the rents generated by the investment (with an adjustment for the depreciation of capital). This present value is then solved to yield an expression for the user cost of capital, $r = (i+\delta)P^I - \Delta P^I$. The problem, then, is to find a way of measuring r or its components. Direct estimates of the user cost are available for only a small fraction of the universe of capital goods (those that are rented). The alternative is to estimate the components of r. The investment good price, P^I, can be obtained from national accounts data, the depreciation rate, δ, can be taken from the Hulten-Wykoff (1981) depreciation study. The rate of return, i, can be estimated in two ways. First, it can be estimated independently from interest rate of equity return data. This is somewhat problematic because of

multiplicity of candidates, and the need to pick a rate that reflects the risk and opportunity cost of the capital good. Jorgenson and Griliches suggest a second way: impose constant returns to scale and find the implied i that causes the accounting equation $pQ = wL + rK$ to hold.[8] It is only at this point that constant returns is absolutely required for the measurement of TFP.

The quantity of capital, K_t, and the quantity of new investment, I_t, are connected (in this framework) by the perpetual inventory method, in which the stock is the sum of past investments adjusted for deterioration and retirement. The resulting concept of capital is thus defined deterioration, in contrast with undeteriorated "gross" stock concept used in some studies.

On the other hand, Jorgenson and Griliches recognized that output must be measured gross of depreciation if it is to conform to accounting system implied by the strict logic of production theory. This put them in conflict with Denison, who advocated a concept of output net of depreciation, and Solow, who used gross output in his empirical work but preferred net output on the theoretical grounds that it is a better measure of the improvement in the welfare arising from technical progress. The debate over this point with Denison thus seemed to pivot on the research objective of the study, not on technical grounds. However, as we have seen above, the Potential Function Theorem, published after the 1967 Jorgenson and Griliches study, links their gross output approach to the A_t of conventional production theory, implying that the competing views of output cannot be simultaneously true (except in very special cases).

---

[8] The implied value of i is then $[P^C C+P^I I-\delta P^I K-\Delta P^I K]/P^I K$. When there are several types of capital goods, a different δ and P^I is estimates for each type, but arbitrage is assumed to lead to a common i for all assets. Hall and Jorgenson (1967) extended the user cost model to include parameters of the income tax system.

Another major contribution of the Jorgenson-Griliches study was to disaggregate capital and labor into their component parts, thereby avoiding the aggregation bias associated with internal shifts in the composition of the inputs (e.g., the compositional bias due to a shift from long-lived structures to shorter-lived equipment in the capital stock, or the bias due to the shift toward a more educated work force). The Divisia index framework was applied consistently to the aggregation of the individual types of capital and labor into the corresponding subaggregate, and applied again to arrive at the formulation in (7). However, because data are not continuous over time but come in discrete-time units, Jorgenson and Griliches introduced a discrete-time approximation to the Divisia derived from the Tornqvist index.[9]

In sum, Jorgenson and Griliches tied data development, growth accounting, and production theory firmly together. The three are mutually dependent, not an ascending hierarchy, as is commonly supposed. These linkages were developed further by Christensen and Jorgenson (1969, 1970), who developed an entire income, product, and wealth accounting system based on the mutuality principle.

F. Diewert's Exact and Superlative Index Numbers

The continuous-time theory of the residual developed by Solow provides a simple, yet elegant, framework for productivity measurement. Unfortunately,

---

[9] In the Tornqvist approximation, the continuous-time income shares $s^K{}_t$ and $s^L{}_t$ in (7) are replaced by the average between-period shares. Capital's discrete-time income share is (sKt+sKt-1)/2. Continuous-time growth rates are also replaced with differences in the natural logarithm of the variable. The growth rate of capital, for example, is $\ln(K_t) - \ln(K_{t-1})$.

data do not come in a time-continuous form. One solution, noted above, is to find reasonable discrete-time approximations to the continuous-time model. In this approach, the choice among competing approximation methods is based largely on computational expediency, with the implication that the discrete-time approximation is not derived as an organic part of the theory, thereby weakening the link between theory and measurement.

Herein lies the contribution of Diewert (1976). He showed that the Tornqvist approximation to the Divisia index used by Jorgenson and Griliches was an exact index number if the production function (4) had the translog form developed by Christensen, Jorgenson, and Lau (1973). In other words, the Tornqvist index was not an approximation at all, but exact under the right conditions. Moreover, since the translog production function could also be regarded as a good "second order" approximation to other production functions, the discrete-time Tornqvist index was a sensible choice even if the "world" was not translog. In this event, the degree of "exactness" in the index number depends on the closeness of the translog function to the true production function. Diewert used the term "superlative" to characterize this aspect of the index.

What Diewert showed, in effect, was that the translog specification of the production function served as a potential function for the discrete Tornqvist index in the same way that the continuous production function served as a potential function for the continuous Divisia index.

One important consequence of this result is that the index number approach of the Solow residual is not entirely non-parametric. There is a parametric production function underlying the method of approximation if the discrete-time index is to be an exact measure of Hicksian efficiency.

However, the values of the "inessential" parameters of the translog, i.e., those other than the Hicksian efficiency parameter, need not be estimated if the Solow residual is used.

G. Dispelling the "Measure of Our Ignorance" with Econometrics

If a specific functional form of the technology must be assumed in order to obtain an exact estimate of the efficiency parameter, why not go ahead and estimate all the parameters of that function using econometric techniques? That is, why not estimate the translog relation between Q_t, K_t, L_t, and A_t directly? For one thing, this avoids the need to impose the marginal productivity conditions of the index number approach.[10] Moreover, it gives a full representation of the technology: all the parameters, not just the efficiency term; and every possible path, not just the path actually followed. Moreover, non-competitive pricing behavior, non-constant returns, and factor-augmenting technical change can be accommodated, and embellishments like cost-of-adjustment parameters can be incorporated into the analysis to help "explain" the residual. Why settle for less when so much more can be obtained under assumptions that must be made anyway, e.g., that the production function has a particular functional form like the translog?

The answers to these questions are familiar to practitioners of the productivity art. There are pitfalls in the econometric approach, just as there are with non-parametric procedures. For example, estimation of the

[10] The marginal productivity conditions can be avoided in the direct estimation of the production function. However, the marginal productivity conditions are used in the estimation of the "dual" cost and profit functions that form an essential part of the productivity econometrician's tool kit.

translog (or other flexible) function can lead to parameter estimates that imply oddly shaped isoquants, causing practitioners to place <u>a priori</u> restrictions on the values of these parameters. There is often a question about the robustness of the resulting parameter estimates against alternative ways of imposing restrictions. Even with these restrictions, the abundance of parameters can press on the number of data observations, requiring further restrictions. Additionally, there is the question of the econometric procedures used to obtain the estimates. The highly complicated structure of the flexible models usually requires non-linear estimation techniques, which are valid only under special assumptions, and there are questions about the statistical properties of the resulting estimates. Finally, because the capital and labor variables on the right-hand side of the regression depends in part on the output variable on the left-hand side, there is the danger of simultaneous equations bias.

In other words, the benefits of the parametric approach are purchased at a cost. It is pointless to debate whether benefits outweigh those costs, simply because there is no reason why the two approaches should be viewed as competitors. In the first place, the output and input data used in the econometric approach are almost always index numbers themselves (there are simply too many types of output and input to estimate separately). Thus, the question of whether or when to use econometrics to measure productivity change is really a question of which stage of the analysis index number procedures should be abandoned. Secondly, there is no reason for an "either-or" choice. Both approaches can be implemented simultaneously, thereby exploiting the relative simplicity and transparency of the non-parametric estimates to serve as a benchmark for interpreting the more complicated results of the parametric

approach. The joint approach has the added advantage that it forces the analyst to summarize the parameters of the translog (or other) function in a way that illuminates their significance for total factor productivity growth (i.e., for the dichotomy between the shift in the production function and factor-driven movements along the function).

There is also a synergism between the parametric and non-parametric approaches. By merging the two approaches, econometrics can be used to disaggregate the total factor productivity residual into terms corresponding to increasing returns to scale, the cost of adjusting the factor inputs, technical innovation, an unclassified trend productivity, and measurement error. Denny, Fuss, and Waverman (1981) were the first to start down this path, and it has grown in importance in recent years. The power of this approach is illustrated by the 1981 paper of Prucha and Nadiri on the U.S. electrical machinery industry. Their version of the total factor productivity residual grew at an average annual rate of 1.99 percent in this industry over the period 1960 to 1980. Of this amount, 35 percent was attributed to technical innovations, 42 percent to scale economies, 21 percent to adjustment cost factors, with only 2 percent left unexplained.

This development addresses the measure-of-our-ignorance problem posed by Abramovitz. It also provides a theoretically rigorous alternative to Denison, who attempted to explain the residual with informed guesses and assumptions which were above and beyond the procedures used to construct his estimates of the residual. And, it speaks to the Jorgenson and Griliches hypothesis that the residual ought to vanish if all explanatory factors could be measured.

H. Digression on Research and Development Expenditures

Another contribution made by Jorgenson and Griliches (1967) was their recognition that aggregate measures of capital and labor included the inputs used in research and development programs to generate technical innovations. Thus, some part of the rate of innovation that drove the TFP residual was already accounted for in the data. As a result, if the social rate of return to the R&D expenditures buried in the input data is equal to the private return, the effect of R&D would be fully accounting for and innovation component of the residual should disappear. On the other hand, if there is a wedge between the social and private rates of return, the innovation component of the residual should reflect the externality. This is a harbinger of the New Growth Theory view of endogenous technical innovation.

The important task of incorporating R&D expenditures explicitly into the growth accounting framework has, unfortunately, met with limited success. Griliches (1973) pointed out a key problem: direct R&D spending is essentially an internal investment to the firm, with no observable "asset" price associated with the investment "good" nor observable income stream associated with the stock of R&D capital. As a result, there is no ready estimate of the quantity of knowledge capital or its growth rate, nor of the corresponding share-weight, which are needed to construct a Divisia index. Moreover, much of the R&D effort of any private firm goes to improving the quality of the firm's products, not the productivity of its production process (more on this later).

There is, of course, a huge literature on R&D and the structure of production, but it is almost entirely an econometric literature (see Nadiri

(1993) and Griliches (1994) for reviews). A satisfactory account of this literature is well beyond the scope of a biography of the non-parametric residual.

I. The Comparison of Productivity Levels

The total factor productivity residual defined above is expressed as a rate of growth. The growth rate of TFP is of interest for intertemporal comparisons of productivity for a given country or region at different points in time, but it is far less useful for comparing the relative productivity of different countries or regions. A developing country may, for example, have a much more rapid growth of total factor productivity than a developed country, but start from a much lower level. Indeed, a developing country may have a more rapid growth of total factor productivity than a developed country <u>because</u> it starts from a lower level and is able to import technology. This possibility is discussed in the huge literature on "convergence theory."

The first translog non-parametric estimates of TFP levels were developed by Jorgenson and Nishimizu (1978) for the comparison of two countries. This innovation was followed by an extension of the framework to include the comparison of several countries simultaneously by Caves, Christensen, and Diewert (1982a)). Moreover, in a contemporaneous paper, Caves, Christensen, and Diewert (1982b) apply a different approach -- the Malmquist index -- to the comparison of relative productivity levels.

The Malmquist index asks the simple question: how much output could country A produce if it used country B's technology with its own inputs. Then: how much output could country B produce if it used country A's technology with its inputs. The Malmquist productivity index is the geometric

means of the answers to these two questions. If, for example, the output of country A would be cut in half if it were forced to use the other country's technology, while output in country B would double, the Malmquist index would show that A's technology is twice as productive.[11] When the production functions differ only by the Hicks-neutral efficiency index, A_A and A_B, respectively, the Malmquist index gives the ratio A_A/A_B. This is essentially the Solow result in a different guise. Moreover, when the technology has the translog form, Caves, Christensen, and Diewert (1982b) show that the Tornqvist and Malmquist approach yield the same result.

However, the two approaches may differ if efficiency differences are not Hicks-neutral, or if there are increasing returns to scale. In these situations, the relative level of technical efficiency will depend on the input levels at which the comparison is made. If, by some chance, other input levels had happened to occur, the Malmquist index would have registered a different value, even though the production functions in countries A and B were unchanged. This is the essence of the path dependence problem in index number theory.

Malmquist indexes have been used in productivity measurement mainly in the context of non-parametric frontier analysis (e.g., Färe et. al. (1994)). Frontier analysis is based on the notion of a best-practice level of technical efficiency which cannot be exceeded, and which might not be attained. An

---

[11] Formally, let $Q_A = F(X_A)$ be the production function in country A and $Q_B = G(X_B)$ in country B. The Malmquist approach estimates how much output Q^*_A would have been produced in A if the technology of B had been applied to A's inputs, i.e., $Q^*_A = G(X_A)$. The ratio Q_A/Q^*_A is then a measure of how much more (or less) productive is technology A compared to technology B at A's input level. A similar calculation establishes the ratio Q^*_B/Q_B, which measures how much more productive technology B is when compared to that of A at the input level prevailing in country B. The Malmquist index is the geometric mean of the two ratios.

economy (or industry or firm) may be below its best-practice level for a variety of reasons: obsolete technology, poor management, constraints on the use of resources, and so on. A measured change in the level of efficiency may therefore reflect an improvement in the best-practice technology or an improvement in the management of the prevailing technology. Sorting out which is which is an important problem in productivity analysis.

Frontier analysis tackles this problem by using linear programming techniques to "envelope" the data and thereby locate the best-practice frontier. The main advantages of frontier analysis are worth emphasizing. First, frontier techniques allow the observed change in total factor productivity to be resolved into changes in the best-practice frontier and changes in the degree of inefficiency. Second, the technique is particularly useful when there are multiple outputs, some of whose prices cannot be observed (as when, for example, negative externalities like pollution are produced jointly with output). The principle drawback arises from the possibility that measurement errors may lead to data which are located beyond the true best-practice frontier. These outliers will be "enveloped" mistakenly by frontier techniques, resulting in an erroneous best-practice frontier.

J. Capital Stocks and Capacity Utilization

Production functions are normally defined as a relation between the flow of output, on the one hand, and the flows of capital and labor services on the other. If the residual is to be interpreted as the shift in an aggregate production function, the associated variables must be measured as flows. This

is not a problem for output and labor, since annual price and quantity data are available. Nor would it be a problem for capital goods if they were rented on an annual basis, in which case there would be little reason to distinguish them from labor input. Capital goods are, however, most often used by their owners. Thus, we typically observe additions to the stock of goods, but not the stock itself nor the services flowing from the stock. Stocks can be imputed using the perpetual inventory method (the sum of net additions to the stock), but there is no obvious way of getting at the corresponding flow of services.

This would not be a problem if service flows were always proportional to the stock, but proportionality is not a realistic assumption. As economic activity fluctuates over the business cycle, periods of high demand alternate with downturns in demand. Capital stocks are hard to adjust rapidly, so periods of low demand are typically periods of low capital utilization, etc. A residual calculated using capital stock data thus fluctuates procyclically along with the rate of utilization. These fluctuations tend to obscure the movements in the longer-run components of the residual and make it hard to distinguish significant breaks in trend. The dating and analysis of the productivity slowdown of the 1970s is an important case in point.

Jorgenson and Griliches address this problem by adjusting capital stock for a measure of utilization based on fluctuations in electricity use. The form of this adjustment became part of the controversy with Denison, but the real problem lay with the use of <u>any</u> externally imposed measure of capital utilization. Any such measure leads to the a theoretical problem: how does a direct measure of capital utilization enter the imputed user cost? Indeed, shouldn't the opportunity cost of unutilized capital be zero?

Berndt and Fuss (1986) provide an answer to these questions. They adopt the Marshallian view that capital stock is a quasi-fixed input in the short-run, whose income is the residual left over after paying off the current account inputs. In terms of the fundamental accounting identity, the residual return to capital is $rK = pQ-wL$, where K is the stock of capital (not the flow) and r is the *ex post* cost of using the stock for one period. Fluctuations in demand over the business cycle cause *ex post* returns to rise or fall relative to the *ex ante* user cost on which the original investment was based. The key result of Berndt and Fuss is that the *ex post* user cost equals the actual (short run) marginal product of capital, and is thus appropriate for use in computing the TFP residual. Moreover, since the *ex post* user cost already takes into account fluctuations in demand, no separate adjustment is, in principle, necessary.

One the negative side, it must be recognized that the Berndt-Fuss revisions to the original Solow residual model fail, in practice, to remove the procyclical component of the residual. This failure may arise because the amended framework does not allow for the entry and exit of firms over the business cycle (and, indeed, is only a partial theory of capital adjustment). Additionally, this approach to utilization does not generalize to multiple capital goods. However, the Berndt-Fuss insight into the nature of capital utilization, and its relation to the marginal product of capital, is a major contribution to productivity theory: it clarifies the nature of capital input and illustrates the *ad hoc* and potentially inconsistent nature of externally imposed utilization adjustments.[12]

[12] The "dual" approach to the Berndt-Fuss utilization model is explored in Hulten (1986). This papers clarifies the links between average cost, TFP, and the degree of utilization.

III. Recent Developments and the Paths Not Taken

The 1980s were a high-water mark for the prestige of the residual, and a watershed for non-parametric productivity analysis as a whole. The Bureau of Labor Statistics began publishing "multi-factor" productivity (their name for total factor productivity) estimates in 1983, and major contributions continued outside the government, with the articles already noted, and the books by Jorgenson, Gollop, and Fraumeni (1987), and Baumol, Blackman, and Wolff (1989). There has also been an interest in applying growth accounting to explain international differences in growth (e.g., Dowrick and Nguyen (1989), and the controversy triggered by Young (1992,1995), and literature on infrastructure investment inspired by Aschauer (1989)). However, the tide had begun to turn against the aggregative non-parametric approach pioneered by Solow, Kendrick, Jorgenson and Griliches, and Denison. Several general trends are discernible:

- The growing preference for econometric modeling of the factors causing productivity change.
- The shift in attention from the study of productivity at the aggregate and industry level of detail to the firm and plant level.
- A shift in emphasis from the competitive model of industrial organization to non-competitive models.
- The effort to incorporate firm-level research and development and patenting into the explanation of productivity change, and
- A growing awareness that improvements in product quality were potentially as important as process-oriented innovation that improved the productivity of capital and labor.

There were several reasons for this shift in focus. The explosion in

computing power enabled researchers to assemble and analyze large sets of data. High-powered computers are so much a part of the current environment that it is hard to remember that much of the seminal empirical work done in the 1950s and early 1960s was done by hand or on mechanical calculating machines (or, later on, by early main-frame computers that were primitive by today's standards). Anyone who has inverted a five-by-five matrix "by hand" will know why multivariate regressions were not often undertaken. The growth of computing power permitted the estimation of more sophisticated, multi-parametered production and cost functions (like the translog), and created a derived demand for large data sets like the Census' LRD, which came into play in 1982.

The arrival of the "New Growth Theory" was a more evident factor behind the shift in the research agenda of productivity analysis. New Growth Theory challenged the constant-returns and perfect-competition assumptions of the total factor productivity residual by offering a view of the world in which (a) markets were non-competitive, (b) the production function exhibited increasing returns to scale, (c) externalities among micro-units were important, and (d) innovation was an endogenous part of the economic system. This shift in perspective gave an added push to the investigation of micro data sets and to the interest in R&D as an endogenous explanation of output growth.

These factors would have sufficient to redirect the research agenda of productivity analysis. However, it was the slowdown in productivity growth, which started sometime between the late 1960s and the 1973 OPEC oil crisis, that settled the matter. Or, more accurately, the failure of conventional productivity methods to provide a generally accepted explanation for the

slowdown virtually guaranteed that the assumptions of the conventional analysis would be changed and explanations sought elsewhere.[13] The residual was, after all, still the "measure of our ignorance," and the New Growth paradigm and the large-scale micro-productivity data sets arrived just in time to fill the demand for their existence.

The directions taken by productivity analysis in recent years are not easy to summarize in a unified way. I will, however, offer some comments on recent developments in the field in the remaining sections. They reflect, to some extent, my own research interests and knowledge, and make no pretense of being an exhaustive survey.

IV. Productivity in the Context of Macro-Growth Models

A. The "Old" Growth Theory

The total factor productivity model produces an explanation of economic growth based solely on the production function and the marginal productivity conditions. Thus, it is not a theory of economic growth because it does not explain how the right-hand side variables of the production function -- labor, capital, and technology -- evolve over time. However, Solow himself provided an account of this evolution in a separate and slightly earlier paper (1956). He assumed that labor and technology were exogenous factors

[13] The literature on the productivity slowdown is voluminous, and still growing (see, for examples, Denison (1979), Berndt (1980), Griliches (1980), Maddison (1987), Diewert and Fox (1999)). Many different explanations have been offered, from the failure to measure output correctly (particularly in the service sector) to the measurement problems posed by high rates of inflation. No single explanation has won out, nor has a consensus emerged about the relative importance of the various competing alternatives.

determined outside the model, and that investment is a constant fraction of output. Then, if technical change is entirely labor-augmenting and the production function is well-behaved, the economy converges to a steady-state growth path along which output per worker and capital per worker grow at the rate of technical change. Cass (1965) and Koopmans (1965) arrive at essentially the same conclusion using a different assumption about the saving-investment process.

Both of these "neoclassical" growth models produce a very different conclusion than the TFP model about the importance of technical change as a cause of economic growth. In the neoclassical growth models, capital formation explains none of the long-run steady-state growth in output, because capital is itself endogenous and driven by technical change: technical innovation causes output to increase, which increases investment, which thereby induces an expansion in the stock of capital. This induced capital accumulation is the direct result of TFP growth and, in steady-state growth, all capital accumulation and output growth are due to TFP. While real-world economies rarely meet the conditions for steady-state growth, the induced-accumulation effect is present outside of steady-state, whenever the output effects of TFP growth generates a stream of new investment.

What does this mean for the measurement of TFP? The residual is a valid measure of the shift in the production function under the Solow assumptions. However, because the TFP residual model treats all capital formation as a wholly exogenous explanatory factor, it tends to overstate the role of capital and understate the role of innovation in the growth process.[14] Since some

[14] This was pointed out in Hulten (1975,1978) in the context of the neoclassical model, and by Rymes (1971) and Cas and Rymes (1991) in a somewhat different context.

part of the observed rate of capital accumulation is a TFP-induced effect, it should be counted along with TFP in any assessment of the impact of innovation on economic growth. Only the fraction of capital accumulation arising from the underlying propensity to invest at a constant rate of TFP growth should be scored as capital's independent contribution to output growth.[15]

The distinction between the size of the residual, on the one hand, and its impact on growth, on the other, has been generally ignored in the productivity literature. This neglect has come back to haunt the debate over "assimilation versus accumulation" as the driving force in economic development. A number of comparative growth studies have found that the great success of the East Asian Tigers was driven mainly by the increase in capital and labor rather than by TFP growth (Young (1992,1995), Kim and Lau (1994), Nadiri and Kim (1996), and Collins and Bosworth (1996)). With diminishing marginal returns to capital, the dominant role of capital implies that the East Asian Miracle is not sustainable and must ultimately wind down (Krugman (1994)). However, these conclusions do not take into account the induced capital accumulation effect. The role played by TFP growth (assimilation) is actually larger and the saving/investment effect proportionately smaller.

---

[15] This point can be illustrated by the following example. Suppose that an economy is on a steady-state growth path with a Harrod-neutral rate of technical change of 0.06 percent per year. If capital's income share is 1/3 of GDP, a conventional TFP sources-of-growth table would record the growth rate of output per worker as 0.06, and allocate 0.02 to capital per worker and 0.04 to TFP. Observed capital formation seems to explain one-third of the growth in output per worker. However, its true contribution is zero in steady-state growth. The 0.06 growth rate of Q/L should be allocated in the following way: 0 to capital per worker and 0.06 to technical change.

A more complicated situation arises when technical change is also embodied in the design of new capital. In this case, the rate of investment affects the rate of technical change and creates a two-way interaction with TFP growth.

Exactly how much larger is hard to say, because the induced-accumulation effect depends on several factors, such as the bias in technical change and the elasticity of substitution between capital and labor. I proposed a correction for this effect in my 1975 paper and estimated that the conventional TFP residual accounted for 34 percent of U.S. output growth over the period 1948 to 1966 (annual output growth was 4.15 percent and the residual was 1.42 percent). When the induced capital accumulation effect formation was taken into account, technical change was actually responsible for 64 percent of the growth in output. This is almost double the conventional view of the importance of TFP growth.

A closely related alternative is to use the a Harrod-Rymes variant of the TFP residual instead of the conventional Hicksian approach. The Harrodian concept of TFP measures the shift in the production function along a constant capital-output ratio, instead of the constant capital-labor ratio of the conventional Hicks-Solow measure (A_t) of the preceding sections. By holding the capital-output ratio constant when costless innovation occurs, the Harrodian measure attributes part of the observed growth rate of capital to the shift in the production function. Only capital accumulation in excess of the growth rate of output is counted as an independent impetus to output growth. The Harrodian approach thus allows for the induced-accumulation effect, and when the innovation happens to be of the Harrod-neutral form, the accounting is exact (Hulten (1975)). Otherwise, the Harrodian correction is approximate.

When applied to the East Asian economies studied by Young, the Harrodian correction gives a very different view of the role of TFP growth (Hulten and Srinivasan (1999)). Conventional Hicksian TFP accounts for approximately one-

third of output growth in Hong Kong, South Korea, and Taiwan over the period 1966-1990/91. With Harrodian TFP, this figure rises to nearly 50 percent. Again, while the conventional Hicksian TFP residual is a valid measure of the shift in the production function, a distinction must be made between the magnitude of the shift and its importance for output growth.

B. The New Growth Models

Neoclassical growth models assume that innovation is an exogenous process, with the implication that investments in R&D have no systematic and predictable effect on output growth. But, can it really be true that the huge amount of R&D investment made in recent years was undertaken without any expectation of gain? A more plausible approach is to abandon the assumption that the innovation is exogenous to the economic system, and recognize that some part of innovation is, in fact, a form of capital accumulation.

This is precisely the view incorporated in the "endogenous" growth theory of Romer (1986) and Lucas (1988). The concept of capital is expanded to include knowledge and human capital, and added to conventional fixed capital, thus arriving at total capital. Increments to knowledge are put on an equal footing with all other forms of investment, and therefore the rate of innovation is endogenous to the model. The key point of endogenous growth theory is not, however, that R&D and human capital are important determinants of output growth. What is new in endogenous growth theory is the assumption that the marginal product of (generalized) capital is constant, rather than diminishing as in the neoclassical theories. It is the diminishing marginal returns to capital that brings about convergence to steady-state growth in the

neoclassical theory and, conversely, it is constant marginal returns which causes the induced accumulation effect on capital to go on ad infinitum.[16]

Endogenous growth theory encompasses a variety of different models. We will focus here on one of which is, perhaps, the main variant in order to illustrate the implications of the endogenous for the measurement and interpretation of the productivity residual. Suppose that the production function (4) has the Cobb-Douglas production function prevalent in that literature, and that (generalized) capital has two effects: each one percent increase in capital raises the output of its owner-users by β percent, but also spillovers to other users, raising their output by a collective α percent. Suppose, also, that $\alpha+\beta=1$, implying constant returns to scale in the capital variable across all producers, while labor and "private" capital are also subject to constant returns ($\beta+\gamma=1$). This leads to:

$$(8) \qquad Q_t = A_0 K_t^{\alpha} [K_t^{\beta} L_t^{\gamma}], \quad \alpha+\beta = 1, \quad \beta+\gamma = 1 \ .$$

This production function exhibits increasing returns to scale overall, but it is consistent with equilibrium since each producer operated under the assumption of constant returns to the inputs which he controls.

What does this new formulation imply for the residual, computed as per the "usual" equation (7)? The residual is derived from the Hicksian production function (4), and the formulation above is a special case of this function in which the output elasticities are constant (Cobb-Douglas) and the efficiency term $A_0 K^{\alpha}_t$ replaces the Hicksian efficiency parameter A_t. The

---

[16] Barro and Sala-i-Martin (1995) provide a good overview of the various growth models. See also Easterly (1995).

associated residual, analogous to (7), is thus equal to the growth rate of capital weighted by the spillover effect. The "endogenous" TFP residual continues to measure costless gains to society -- the "Manna from Heaven" -- from innovation. But now this Manna is associated with the externality parameter α rather than the Hicksian efficiency parameter A_t. Thus, in the New Growth view, the residual is no longer a non-parametric method for estimating a fixed parameter of the production function, but the reflection of a process. Moreover, there is no reason for the residual to disappear.[17]

The endogenous growth residual adds structure to the problem of measuring TFP, but is it superior to the old-fashioned variety in addressing issues like the productivity slowdown? The old approach came under attack because it failed to explain the productivity slowdown. Does the new view do any better? The endogenous growth view, in the form set out above, points either to a slowdown in the growth rate of (comprehensive) capital or to a decline in the degree of the externality α as possible causes of the slowdown. Unfortunately, neither possibility is supported by the available evidence. Investment in R&D as a percent of GDP has been relatively constant and the proportion of industrial R&D has increased. The growth in fixed capital does not correlate with the fall in the residual. Moreover, the evidence does not provide support for a decline in the externality or "spillover" effect (Nadiri (1993), Griliches (1994)), although this is debatable. It therefore seems

[17] These conclusions assume that the spillover externality augments the return to labor and "private" capital equally (an implication of the Cobb-Douglas form). All is well if labor continues to be paid the value of its marginal product. However, endogenous growth theory is part of a more general view of growth that stress the importance of imperfect competition, and it is possible that the presence of spillover externalities may lead to a wedge between output elasticities and factor shares.

fair to say that, at this point, endogenous growth theory does not offer a clear way out of the slowdown conundrum.

C. Data on Quality and the Quality of Data

Although it has not vanquished its older rival in explaining the productivity slowdown, the New Growth Theory has focused attention on the nature of innovation and offered a better conceptual explanation than the older growth literature, which was largely content to associate innovation with the unexplained shift in the production function. However, the endogenous growth model and exogenous Hicks-neutral model both presume that technical change is <u>process-oriented</u>, improving the technological process by which input is transformed into output. Both approaches thus ignore another important dimension of innovation: improvements in the quality of products and the introduction of new goods. Both present consumers and producers with a new array of products and, over time, the market basket is almost completely transformed (automobiles replace horses, personal computers replace typewriters, etc.). Much of the welfare gain from innovation comes from the production of better goods, and not just from the production of more goods. Unfortunately, the TFP residual is intended to measure only the production of more goods -- this is what a shift in the production function means -- and only the costless portion at that. Innovation that results in better goods is not part of the TFP story.

One way to handle this issue is to treat the two types of innovation as separate measurement problems and restrict the use the TFP residual to its

proper domain. Unfortunately, the two types of innovation are not easily segregated, as the following example shows. First, imagine that there are two economies, both of which start with 100 units of input and both of which have the same technology, so that both produce 100 physical units of output. Suppose, now, that some ingenious person in economy A discovers a way to double the amount of output that the 100 units of input can produce. At the same time, an innovator in economy B discovers a way to double the utility of the 100 physical units of output that are produced (that is, inhabitants of B gladly exchange two units of the "old" output for one unit of "new" output). Measured TFP will double in A but remain flat in B, even though the inhabitants of both countries are equally well off as a result of their respective innovations.

Is this the right result? In a sense, it is. The production function for physical units of output shifted in economy A but not in B. However, this judgment reflects a particular conception of output, i.e., that physical units are the appropriate unit of measure. This convention obviously provides an unfavorable view of economy B, since it defines away the true gains made in B. An alternative approach would be to measure output in units of consumption efficiency, that is, in units that reflect the marginal rate of substitution between old and new goods. In this efficiency-unit approach, both A and B experience a doubling of output, albeit for different reasons, and measured TFP reflects the increase. In other words, the TFP model does service in measuring both process and product innovation when output is measured in efficiency units.

The efficiency approach to productivity measurement has proceed along two general lines. First, the 1950s saw the theoretical development of the

model of capital-embodied technical change (Johansen (1959), Salter (1960), Solow (1960)). In this model, technical innovation is expressed in the design of new machines, with the implication that different vintages of capital may be in service with different degrees of marginal productivity. When expressed in efficiency units, one physical unit of new capital represents more capital than one physical unit of an older vintage. The total "size" of the capital stock is the number of efficiency units it embodies.

While theoretically plausible, the capital-embodiment model met initially with limited empirical success. Moreover, it was dismissed as "unimportant" by one of the leading productivity analysts, Denison (1964). However, the issue did not disappear entirely, and has returned to prominence with hedonic price study by Cole et. al. (1986), who used price data to show that official investment-price statistics had essentially missed the computer revolution, overstating price and understating quantity (measured in efficiency units).[18] This finding led BEA to switch to an efficiency-unit convention for investment in computers in the U.S. national income and product accounts (but only for computers). This analysis was extended by Gordon (1990), who adjusted the prices of a wide range of consumer and producer equipment for changes in quality. Gordon also found systematic overstatement of official price statistics and a corresponding understatement of efficiency-

[18] In the hedonic-price model, a product is viewed as a bundle of constituent characteristics. The more there is of each characteristic, the more there is of the good. Computers, for example, are seen in terms of CPU speed, memory speed and capacity, storage capacity, etc. The hedonic model estimates a "price" for each characteristic, and thereby derives an implied price for the whole bundle. This also yields a "quantity" of the good measured in efficiency unites. Embodied technical change is naturally seen as an increase in the efficiency units via an increase in the characteristics. See Triplett (1983,1987) for more on the hedonic-price model.

adjusted quantity investment output and the resulting capital input.

The Consumers Price Index is another area in which price data are routinely adjusted for "quality" change. A variety of procedures are used in adjustment process, including price hedonics, but the Advisory Commission (1996) concluded that they were not adequate and that the CPI was biased upward by 0.6 percentage points per year. In other words, the growth in efficiency-price of consumption goods was overstated and the corresponding quantity understated. The Bureau of Labor Statistics is currently undertaking revisions in its procedures, including increased reliance of price hedonics, to address the quality problem.

The fundamental problem with efficiency approach is that improvements in product quality, or the advent of entirely new consumer goods, are essentially subjective. Physical units can be observed, however imperfectly, but when characteristic/efficiency units are involved, there is no direct observational check to the imputed amount of product. It is all too easy to misstate the true quantity of efficiency units, and there little intuitive basis for rejecting the misstatement.[19] It is worth recalling the words of Adam Smith,

---

[19] The mismeasurement of quality in improved products is particularly difficult in regard to nondurable consumer goods, where overlapping prices between old and new models are either absent or unreliable. Moreover, the measurement problems posed by "quality" are not limited to product-oriented innovation. There are also myriad problems in the definition of output which involve a quality dimension without reference to innovation. Griliches (1994) speaks of the "hard to measure" sectors of the economy, largely the service sector, and notes that it is precisely these sectors that have grown over time. For example, the revenue of banks can be measured with some precision, but what exactly are the units of output? How would one measure these units in principle, and account for differences in the quality of service that are characteristic of competition among banks? Unless the nature of the output can be defined precisely, it is impossible to determine its rate of growth and to confront questions about the impact of quality-enhancing innovations like automatic teller machines.

"Quality ... is so very disputable a matter, that I look upon all information of this kind as somewhat uncertain."

The subjective nature of the efficiency approach leads to a more subtle problem. Since the quantity of efficiency units is determined by imputation of the relative marginal utility between old and new products, the very definition of product quantity becomes a matter of utility and consumer choice. This tends to blur the boundary between the supply-side constraint on growth, the production function, and the objective of growth, which is the province of the utility function. We will return to such boundary issues in the following sections.

D. Quality Change and the Residual

Most of the TFP studies that have incorporated product-oriented innovation into the residual have focused on capital-embodied technical change. Nelson (1964) expressed the residual as a function of the rate of embodiment and the average age of the capital stock. Domar (1963) and Jorgenson (1966) observed that capital is both an input and an output of the production process, and the failure to measure capital in efficiency units causes two types of measurement error: one associated with the mismeasurment of capital input and an error associated with the mismeasurement of investment good output. Surprisingly, the two errors exactly cancel in Golden Rule steady-state growth, leaving the residual unbiased.[20]

---

[20] This point is often overlooked in econometric studies of embodied technical change. If both capital input and investment output are correctly measured in efficiency units, the economy-wide TFP residual should be invariant to changes in the rate of capital embodiment. If input and output is not adjusted for quality, aggregate

The actual size of the input and output embodiment errors depends on the rate at which embodied efficiency increases and the average embodied efficiency of the older vintages of capital stock. These cannot be estimated within the residual's index number framework, but in my 1992b paper, I use data from Gordon (1990) to estimate the net embodiment effect for U.S. manufacturing industry. The net embodiment effect was found to account for about 20 percent of the TFP residual over the time period 1949-1983. Wolff (1996) reports an effect which is roughly twice as large for the economy as a whole for the same years. Greenwood et. al. (1997) propose a variant of the embodiment model in which the total value of investment is deflated by the price of consumption rather than investment. The resulting estimate of the embodiment effect accounts for 58 percent of the aggregate residual, per the period 1954-1990.

These studies deal with capital-embodied technical change. Productivity analysis has paid less attention to quality change in consumption goods. The example of the economies A and B from the preceding section suggests that this neglect results in an understatement of true output and TFP growth (recall the situation in economy B). However, the problem is even more complicated than that example suggests, because of another problem that has lurked in the background of productivity analysis: the cost of achieving technical innovations. A variant of our example illustrates the problem. A and B both

---

TFP is still invariant along the optimal growth path. Off the optimal path, there is the Hall (1967) identification problem to reckon with: the exponential part of capital-embodied technical change cannot be distinguished from the equivalent rate of disembodied technical change given price or quantity data on age, vintage, and time. Only deviations from the exponential path can be identified. Finally, it is well to remember that the residual can only measure the costless part of innovation, embodied or otherwise.

start with 100 units of input and the same technology, and produce 100 physical units of output. Economy A now invests half its work force in research, and is able to quadruple the output of the remaining 50 workers. Output and TFP thus double. In economy B, on the other hand, the 50 are diverted to research and they manage to invent a new good that is four times as desirable (that is, inhabitants of B gladly exchange four units of the old output for one unit of new), but only 50 units of physical output are produced. Physical output and TFP fall by half in B, even though innovation has made the inhabitants of B as well off as those in A. The failure to measure output in efficiency units thus gives the appearance of technical regress even though progress has occurred.

These considerations can be parameterized and embedded in the TFP model (Hulten (1996)). Suppose that product-oriented technical change proceeds at a rate θ (essentially the marginal rate of substitution between old goods and new goods of superior quality), and the cost of achieving this rate of quality change is $\mu\theta$. Costless innovation occurs when μ equal zero. In a simplified world in which capital and labor are fixed, it can be shown that the TFP residual falls at the rate $\mu\theta$ when output is measured in physical units, but grows at a rate $(1-\mu)\theta$ when efficiency units are used. In the first case, an increase in the rate of innovation θ will actually cause the residual to decrease, resonating with the New Economy critique that the productivity statistics is the failure to count improvement in product quality.[21]

This model underscores a basic property of the total factor

[21] There is also another possibility. Even if output is correctly measured in quality units, the residual can fall if the rate of innovation θ is pushed beyond its cost-effective optimum. In other words, research "booms" can lower total factor productivity if pushed too far.

productivity residual: the residual measures the costless component of innovation. Under the assumptions of the original Solow derivation, the Hicksian shift parameter A_t is "Manna from Heaven"; in the endogenous growth version of TFP, the residual is related to size of the (costless) externality α; in the current context of product-oriented innovation, the correctly specified residual equals $(1-\mu)\theta$, with $(1-\mu)$ representing the costless component of product-oriented innovation (when μ is zero, the pure Manna case prevails).

E. Capacity Versus Welfare Interpretations of the Residual The Problem of Sustainable Consumption

Once it is recognized that product quality adjustments allow consumer-welfare parameters to creep into the TFP residual, the boundary between the supply-side conception of the residual and the demand-side interpretations is blurred. If welfare considerations are permitted inside one region of the supply-side boundary, and they must be if the quality dimension of output is to make sense, perhaps they should be permitted in other boundary areas, like the net versus gross output controversy, where welfare arguments have also been made. After all, a high rate of real GDP growth, and hence a large gross-output productivity residual, can be sustained in the short-run by depleting unreproducible resources at the expense of long-run welfare. Net output solves this problem by controlling for depreciation and environmental damage, and some believe that it thus provides a more accurate picture of sustainable long-run economic growth. Does it not follow that a separate TFP residual based on net output is the appropriate indicator of the contribution of costless technical innovation to sustainable growth?

The short answer is "no." Changes in social welfare can be shown to depend on the standard gross-output concept of TFP, with no need to define a net-output variant of TFP. The result follows from the optimal growth model studied by Cass (1965) and Koopmans (1965), as augmented by Weitzman (1976), in which the intertemporal utility function $U(C_0, \ldots, C_t)$ is maximized (C_t is the amount of consumption t years from the present time). For present purposes, it is useful to assume that prices are proportional to marginal utilities and express the intertemporal welfare problem as one of maximizing the present value equation

$$W_0 = \sum_{t=0}^{\infty} \frac{p_t C_t}{(1+i)^{t+1}} \quad , \tag{9}$$

subject to the production function $C_t + I_t = A_t F(K_t, L_t)$ and the accumulation condition $K_t = I_t + (1-\delta)K_{t-1}$ (here, we revert to the assumption that Hicksian efficiency and labor growth are exogenously determined). The economic problem of contemporary society, at each point in time, is to determine the optimal division of current output between consumption and investment.

This problem was studied by Weitzman (1976), who demonstrated that the optimal consumption path $\{C^*_t\}$ satisfies the condition $p_t C^*_t + p_t \Delta K^*_t$. But this is really nothing more that the Hicksian definition of income: the maximum amount of output that could be consumed each year without reducing the original amount of capital, or, equivalently, "sustainable" consumption. This is the welfare indicator appropriate for the annualized measurement of increments to consumer welfare.

This welfare indicator of output is not the same as GDP. According to the fundamental accounting identity (1), GDP is equal to the gross payments to

capital and labor (as well as p_tQ_t). With some algebraic manipulation based on the Hall-Jorgenson user cost formula, it can be shown that Hicksian income is equal to net factor income or net national product in nominal prices, which differs from gross output by the amount of depreciation (Hulten (1992a)):

$$(10) \qquad p_tC_t^* + p_t\Delta K_t^* = i_tp_tK_t + w_tL_t < p_tQ_t \ .$$

This identity may encourage some to suppose that net national product (NNP) should be used in productivity analysis instead of GDP, since it is associated with maximum intertemporal welfare. However, the two output concepts are complements, not substitutes. The growth in real GDP indicates the expansion of the supply-side constraint in any year, and the residual computed using real GDP measures the change in the efficiency of production as represented by A_t (the shift in production constraint). The growth in NNP cum Hicksian income reveals the extent to which growth has improved society's welfare. These are separate issues and must be kept separate, and it is important to recognize that the gross-output TFP residual fits into the welfare-maximization problem via the production constraint.

This result does raise the question of how the gross-output residual is related to changes in economic welfare. This is a complicated issue that involves treating capital as an intertemporal intermediate product, and linking labor input and technology directly to the attainable consumption path (Hulten (1979)). If the optimal consumption path $\{C^*_t\}$ is chosen, i.e., the one that maximizes (9), an intertemporal consumption-based residual can be derived which is the weighted sum of the TFP residuals:

(11) $$\Omega_{0,T} = \sum_{t=0}^{T} \omega_t \frac{\dot{A}_t}{A_t} \quad .$$

The individual weights in this expression, ω_t, are the respective annual ratios of GDP to total wealth, W_0. They are the intertemporal counterparts of the Domar weights used to aggregate intermediate inputs.

The $\Omega_{0,T}$ residual indicates the increase in optimal consumption associated with changes in the annual (gross output) TFP residuals. It is not a substitute for these residuals, but a complement. It is clear, once again, that the appropriate welfare-based analysis is separate from, and complementary to, the GDP-based analysis of productive efficiency.

F. The Boundaries of Productivity Analysis

We have seen that the boundary between welfare and capacity is not as straight-forward as one might wish. However, two general boundary principles are clear enough: a distinction must be maintained between ends (welfare improvement) and means (production); a distinction must also be maintained between the impulse to save (i.e., defer consumption) and the impulse to invent (productivity). This section deals with yet another boundary: the line between what should be counted as output and input and what should not. This "comprehensiveness" boundary is central to the debate about the desirability of a "Green GDP" raised by environmentalists and discussed in Nordhaus and Kokkelenberg (1999).

A complete set of economic accounts would include information on the price and quantity of <u>every</u> variable that enters into the production or utility function of every agent in the economy. The required list of

variables would extend far beyond the boundaries of the market economy. Goods produced in the household sector would be an important part of the complete accounting system, including work around the home, leisure, and education. Those public goods produced in the government sector and distributed free of direct charge (or at a price that does not reflect marginal cost) must also be part of the accounts, including national defense, public infrastructure, etc. So must goods held in common for private use (environmental variables like clean air and water, parks, forests and mineral deposits), as well as spillover externalities, such as knowledge and congestion, and so on.

This is an impossibly large order to fill. The boundaries of a complete accounting system would include everything that correlates with the production of goods and services and affects economic welfare. Thus, for example, the effects of urbanization and materialism that are alleged correlates of the modern capitalist system could force their way into the complete accounts on the grounds that the breakdown of welfare-enhancing institutions (like family and religion) are the result of these effects. The boundaries of a complete set of economic accounts may thus be extended to include statistics on crime, drug abuse, divorce, etc.

Boundaries drawn this broadly go far beyond the limits of the current national economic accounts, and probably far beyond the comfort limits of most economists. Current national income accounting practice relies primarily on market transactions to generate data. Market transactions, though flawed and incomplete, do provide an objective yardstick for measuring the volume of economic activity, as well as prices and quantities. Market data are also relatively easy to collect. These benefits are, unfortunately, purchased at a price: the narrow focus on the products which are exchanged for money leads

to the exclusion of many goods whose data are harder to obtain. This, in turn, can lead to a distorted picture of the true production possibilities facing an economy. Productivity, in any of its many forms, is essentially a ratio of output to input and will be affected by the omission of any element of the numerator or denominator.

This dilemma can be illustrated by the following simplified example. Suppose that an industry produces a good Q_t, which it sells at marginal cost in the market place for a price P_t. It produces the good using an input X_t, which it purchases in the factor market for w_t, but also uses a good Z_t which is available without cost to the firm. The item Z_t might be a common good, like clean air, or an externality associated with another agent's behavior (e.g., technical knowledge appropriated from other firms in the industry), or self-constructed capital produced in an earlier year (the firm's stock of technical know-how). In any event, the statistician who looks only at market data will record the accounting identity $P_tQ_t = w_tX_t$, and the analyst will reckon productivity to be Q_t/X_t. The true nature of things is, of course, different. The correct accounting identity is ${P^*}_tQ_t=w_tX_t+\rho_tZ_t$, where P^* is the marginal social cost of the good, as opposed to the private cost, P_t, and ρ_t is the implicit cost to using the "free" input Z_t. The true productivity ratio is $Q_t/F(X_t,Z_t)$. The example could be complicated further by supposing that the firm generates an externality as it produces Q_t.[22]

---

[22] There are many candidates for the role of "significant omitted variable." One in particular deserves mention because of its relation to the productivity of the computer revolution. The advent of computers has allowed firms to reduce the number of employees, often resulting in productivity gains to the firm. But this has often come at the expense of the consumer, who must substitute his/her own time for that of the departed employee. Anyone who has waited "on hold" for a telephone connection to a human voice, or suffered through seemingly interminable menu-driven options, will recognize the problem.

In order for the statistician to "get it right," the variable Z_t must be recognized and measured, and imputations made for the shadow prices P^* and ρ_t. The latter is a particularly hard job. Some imputations can be made using technical procedures like price hedonics, but many must be approached with controversial techniques like "willingness-to-pay" criteria (see, for example, the discussion in Nordhaus and Kokkelenberg (1999)). It is even harder for the statistician to proceed when imputation involves a politically sensitive issue like the preservation of the environment, the public's health, or worker or product safety. Partisans with different points of view often impute vastly different amounts to the value of life or protection of the environment. In these cases, the imputation process is thus likely to reflect partisan agendas as much as the true nature of productivity growth.

Some imputations are made in practice in the national accounts, e.g., owner-occupied housing, and quasi-imputations for government "output" are used. However, the bulk of unpriced goods are not included. This seems the safe path to follow, at least for the time being. While the omission of important variables may limit the generality of conclusions that can be drawn from the productivity statistics, at least the results are not subject to the changing winds of ideology or special interests. Nor is the direction of the "boundary bias" clear.[23]

---

[23] The debate over boundaries has generally failed to recognize that the omission of environmental spillovers from official data does not necessarily mean that they are unnoticed. The public feel their effects whether or not they appear in the data and, indeed, rational citizens should make their own corrections to flawed data. A great deal of pro-environment legislation has been informed by the "biased" statistics, and it is unclear whether fixing the bias would have led to a superior outcome.

V. The Total Factor Productivity Residual for Firms and Industries

A. The View from the Top Down

A total factor productivity residual can, in principle, be computed for every level of economy activity, from the plant floor to the aggregate economy. These residuals are not independent of each other, since, for example, the productivity of a firm reflects the productivity of its component plants. Similarly, industry residuals are related to those of the constituent firms, and productivity in the aggregate economy is determined at the industry level. As a result, productivity at the aggregate level will increase if productivity in each constituent industry rises, or if the market share of the high productivity industry increases (and so on down the aggregation ladder). A complete picture of the industrial dynamics of an economy would include a mutually consistent measure of the TFP residuals at each level in the hierarchy, and of the linkages used to connect levels.

The task of constructing this hierarchy of residuals can be approached from the top down, in a process that can be likened to unpeeling an onion in order to reach lower layers of structure. Domar (1961) was the first to work out the problem of "unpeeling" the TFP residual, and to recognize the complication introduced by the presence of intermediate goods. This complication arises because plants and firms in each sub-layer produce goods and services that are used as inputs in the production processes of the plants and firms. As each layer is unpeeled, the magnitude of these intermediate deliveries gets larger. For example, there are no intermediate goods in the aggregate economy because there is only one "industry" at this level of aggregation and all inter-industry flows cancel out.

However, these inter-industry flows "uncancel" in passing to the one-digit industry level of detail. The iron ore delivered to the steel industry is counted in the gross output of the extractive industries, and counted again as part of the gross output of the manufacturing industry. The sum of the one-digit industry gross output is therefore larger than total aggregate output.

The nature of this problem can be made more precise by observing that the total output of an industry (plant, firm) is composed of deliveries to final demand plus deliveries of the industry's output to the other industries that use the good. On the input side, the firm uses not only labor and capital, but also intermediate goods purchased from other industries. This leads to the following accounting identity:

$$(12) \qquad p_i D_i + p_i \Sigma_i M_{i,j} = w_i L_i + r_i K_i + \Sigma_j p_{j,i} M_{j,i} \ .$$

The summation term on the left-hand side of this expression is the value of the deliveries of the ith industry's output and D_i denotes deliveries to final demand (time subscripts have been omitted for clarity of exposition). The summation on the right-hand side is the value of intermediate goods purchased from other industries, and the remaining terms on the right-hand side constitute the value added by the industry, $w_i L_i + r_i K_i$.

There is an expression like (12) for each industry (firm, etc.) in the economy. Summing them all up to the aggregate level gives the identity:

$$(13) \qquad \Sigma_i p_i D_i = \Sigma_i w L_i + \Sigma_i r K_i = wL + rK \ ,$$

(it is assumed, here, that competition equates wages and capital cost across sectors). This is a variant of the fundamental accounting identity with which we started, but here we have total deliveries to final demand as the output measured on the left-hand side and total value added on the right-hand side.

TFP residuals can be obtained from both expressions -- industry residuals from (12) and the aggregate residual from (13) cum (1). Domar (1961) showed that the aggregate residual is the weighted sum of the industry residuals, where the weights are the ratio of industry gross output to total deliveries to final demand (GDP). His results can be generalized to:

$$\text{(14)} \qquad \frac{\dot{A}_t}{A_t} = \sum_{i=1}^{N} \frac{p_{i,t}\, Q_{i,t}}{\Sigma_i\, p_{i,t}\, D_{i,t}} \frac{\dot{A}_{i,t}}{A_{i,t}} \; .$$

The unusual feature of this expression is that the weights sum to a quantity greater than one to account for the presence of the intermediate goods. Thus, for example, a uniform one percent rate of increase in productivity at the industry level may translate into, say, a one and a half percent increase in productivity at the aggregate level. This inflation in the aggregate number is needed in order to account for the fact that, while an increase in industry-level productivity augments the production of intermediate goods, these intermediate goods have subsequently disappeared in the process of aggregation.[24]

The production function underlying the residual in (14) is the second

[24] It is no accident that (14) looks very much like (11), the welfare equivalent of the Solow residual. The welfare residual is based on the intertemporal optimization of consumption, and capital is treated as an intermediate good in that model. Moreover, "years" are formally equivalent to industries in the conventional intermediate goods model described in this section.

unusual feature of the analysis. Where Solow assumed that the aggregate production function could be expressed as $Q = AF(K,L)$, the technology underlying (14) is a production possibility frontier of the following form: $F(Q_1, \ldots, Q_n; K, L, A_1, \ldots, A_n) = 0$. The left-hand side of (14) is the shift in the frontier holding capital and labor constant. The right-hand side indicates that this shift can be "unpeeled" into separate components: the growth rates of industry-level productivity (the A_i), and the sectoral share-weights, which may change with the reallocation of GDP among sectors with different TFP levels and growth rates. There is no guarantee that the aggregate productivity index is path independent when the component A_i grow at different rates.

The chief difficulty with this unpeeling process lies in the nature of intermediate goods. The quantity gross output and intermediate goods in any industry are greatly affected by mergers and acquisitions. The merger of firms can transform what were once inter-firm flows of goods into intra-firm flows, thereby extinguishing some amount of gross output. This has led some researchers to use real value added, a concept of industry output which is immune to this problem.

The productivity analyst's job would be made easier if intermediate goods could be netted out directly in the identity (12), leaving industry final demand equal to value added, i.e., $p_iD_i = w_iL_i+r_iK_i$. However, this will generally not happen, since the value of intermediate goods produced in an industry need not equal the amount used. One solution is to focus on the right-hand side of this expression and define industry output as the "real," or constant price, part of $w_iL_i+r_iK_i$. Industry value added sums to total value added (GDP), and the relation between the two is not affected by intermediate

goods. A variant of the TFP residual can be based on this concept of industry "output" by applying the original Solow formula. The result can be weighted up to the aggregate level using value added weights.

There are, however, two problems with this approach. First, there is nothing in the real world that resembles real value added. Do plants actually make things in units of real value added? Second, it is well known that the real value added works only when innovation enhances the productivity of capital and labor but not intermediate inputs (i.e., the industry-level production function has the form $Q_i = F[M_i, A_i G(K_i, L_i)]$. Thus, the productivity analyst is confronted with a dilemma: use the gross output approach and become a prisoner of the degree of vertical and horizontal industrial integration, or use the implausible value-added approach. Moreover, there is no guarantee that the production functions underlying either approach are suitable potential functions for the path independent line integration required in (14), and many other problems are encountered at the industry-level of analysis (Gullickson and Harper (1998)).

B. The View from the Bottom Up

The preceding remarks take the "top-down" view of sectoral productivity analysis, in which the aggregate TFP residual is the point of reference. The bottom-up approach to productivity measurement starts from a very different perspective. It takes the universe of plants or firms as the fundamental frame of reference and does not impose the restrictive aggregation assumptions needed to achieve a consistent measure of overall productivity. Instead, the basic heterogeneity of the micro production units is stressed. An important

goal of this approach is to explain the observed heterogeneity of plant productivity in terms of factors like R&D spending or patenting, or differences in the financial or industrial structure.

The literature on this approach is huge and can only be treated with a cursory overview. Early contributions were made by Griliches, Mansfield, and others (see Griliches (1994)), and the work of Nelson and Winter explicitly focused on heterogeneity. This line of investigation was greatly aided by the development of micro data sets like the LRD in 1982 and by the enormous increase in computing power, which enabled the researcher to analyze increasingly large data sets with ever more sophisticated econometric techniques. The R&D work of Griliches and Mairesse (1984) and B. Hall (1993) is noteworthy in this regard, as are the seminal contributions of Davis and Haltiwanger.

The heterogenous plant/firm approach has much to recommend it, since it permits a detailed examination of the factors that actually determine micro productivity. However, its very success is also its greatest weakness: it is hard to generalize the lessons learned from the micro analysis. This is due in part to the inherent heterogeneity of the data, but it is also due to the diverse (and often contradictory) findings of different econometric studies. Moreover, the micro approach to productivity has not succeeded any better than the macro approach in explaining the slowdown in aggregate productivity, nor has it resolved the Solow paradox.

Several studies have attempted to link the micro and macro levels of analysis. Baily, Hulten, and Campbell (1992) used data from the LRD to examine the internal dynamics of industry-level residuals. This study found, among other things, that the representative agent model, which is often

appealed to as the conceptual link between macro and micro levels of analysis, is not supported by the data. When industry-level residuals were resolved into the weighted sum of the plant-level residuals, it was found that the plants with rising levels of TFP and plants with high pre-existing TFP levels were the main contributors to productivity growth. Firms with low pre-existing TFP levels and declining firms were a drag on productivity. The persistence of firms with both high and low levels of productivity suggests a more complex view of industrial organization than the simple representative agent model used to motivate the aggregate TFP residual. The micro data also suggest a more complex productivity dynamic, in which the entry and exit of firms, as well as their expansion and contraction, are important dimensions.

Many advances have been made in subsequent research. However, it remains true that a compelling link between the micro and macro levels has yet to be forged. This is one of the greatest challenges facing productivity analysis today. This challenge is all the more daunting because it must confront this problem: industries are composed of heterogenous firms operated under conditions of imperfect competition, while the theoretical aggregation conditions required to proceed upward to the level of the macro economy rely on perfect competition.

VI. Conclusion

Any respectable biography must end with a summary judgment about the subject at hand and, above all, the true character of the subject should be revealed. This is particularly important in the case of the total factor productivity residual, whose true character has often been misunderstood by friends and critics alike. The portrait painted in this paper reveals these

essential features:

- The residual captures changes in the amount of output that can be produced by a given quantity of inputs. Intuitively, it measures the shift in the production function.

- Many factors may cause this shift: technical innovation, organizational and institutional change, shifts in societal attitudes, fluctuations in demand, changes in factor shares, omitted variables, and measurement errors. The residual should not be equated with technical change, although it often is.

- To the extent that productivity is affected by innovation, it is the costless part of technical change that it captures. This "Manna from Heaven" may reflect spillover externalities thrown off by research projects, or it may simply reflect inspiration and ingenuity.

- The residual is a nonparametric index number designed to estimate one parameter in the larger structure of production, the efficiency shift parameter. It accomplishes this by using prices to estimate marginal products.

- The various factors comprising the TFP are not measured directly, but lumped together as a residual "left over" factor (hence the name). They cannot be sorted out within the pure TFP framework, and this is the source of the famous epithet "measure of our ignorance."

- The Divisia index must be path in dependent in order for uniqueness. The discrete-time counterpart of the Divisia index, the Tornqvist approximation, is an exact index number if the underlying production function has the translog form. The problem of path dependence is a problem of uniqueness, and this is not the same thing as measurement bias.

- The conditions for path independence are (1) the existence of an underlying production function and (2) marginal productivity pricing. Neither constant returns to scale nor Hicks neutrality are absolutely necessary conditions, although they are usually assumed for convenience of measurement.

- When the various assumptions are met, the residual is a valid measure of the shift in the production function. However, it generally understates the importance of productivity change in stimulating the growth of output because the shift in the function generally induces further movements along the function as capital increases.

- The residual is a measure of the shift in the supply-side constraint on welfare improvement, but it is not intended as a direct measure of this improvement. To confuse the two is to confuse the constraint with the objective function.

This is the essential character of our subject. As with any portrait which is examined closely, flaws are detected and the final judgment is usually mixed with praise and criticism.

Had the total factor productivity approach provided a clear explanation of the productivity slowdown, much of the criticism would be confined to academic sniping at minor points of methodology. However, not only did the residual approach fail to provide a convincing explanation of the slowdown, it seemed to indicate that there was little productivity growth at a time when many observers were convinced that a technological revolution was underway. This disconnect between data and New Economy perceptions challenged the credibility of the procedure. Paradoxically, another set of observers were convinced that the same data and procedures were biased in the opposite direction, on the grounds that the productivity numbers tend to omit many of the negative correlates of growth.

A fair judgement must go beyond these criticisms and address a more fundamental question: to what extent are the perceived failures inherent in the character of the residual, and to what extent are the problems inherent in the data to which the residual technique is applied? If data on prices and quantities do not accurately reflect quality improvement, or the boundaries of the data set are drawn to closely, attacking TFP is rather like shooting the messenger because of the message. If the data are the real source of complaint, other methods (e.g., econometrics) will not fair much better than the simple residual. Bad data are bad data regardless of how they are used.

The failure of other techniques to convincingly outperform the TFP residual on important issues may be taken as a limited degree of support, albeit a sort of 'damning with faint praise.' The residual deserves much more

credit. It has provided a simple and internally consistent intellectual framework for organizing data on economic growth, and has provided the theory to guide economic measurement. Moreover, it teaches a lesson that is still not fully appreciated by mainstream economics and national income accounting: that an empirically testable theory places restrictions on the way data must be collected and organized, and that choices about the measurement procedures are often implicit choices about the underlying theory.

The residual is still, after more than forty years, the work horse of empirical growth analysis. For all its flaws, real and imagined, many researchers have used it to gain valuable insights into the process of economic growth. Thousands of pages of research have been published, and the residual has become a closely watched government statistic. Not bad for a forty-year old.

References

Abramovitz, Moses, "Resource and Output Trends in the United States since 1870," American Economic Review, 46, 2, May 1956, 5-23.

Advisory Commission to Study the Consumer Price Index, Toward a More Accurate Measure of the Cost of Living, Final Report to the Senate Finance Committee, Washington D.C., December 4, 1996.

Aschauer, David A., "Is Public Expenditure Productive?," Journal of Monetary Economics, 23, 1989, 177-200.

Baily, Martin N., Charles R. Hulten, and David Campbell, "Productivity Dynamics in Manufacturing Plants," Brookings Papers on Economic Activity, Microeconomics, 1992, 187-249.

Baily, Martin N. and Robert J. Gordon, "Measurement Issues, the Slowdown, and the Explosion of Computer Power," Brookings Papers on Economic Activity, 2, 1988.

Barro, Robert J. and Xavier Sala-i-Martin, Economic Growth, Mc Graw-Hill, New York, 1995.

Baumol, William J., Sue A.B. Blackman, and Edward N. Wolff, Productivity and American Leadership: The Long View, The MIT Press, Cambridge Massachusetts, 1989.

Berndt, Ernest R. and Melvyn A. Fuss, "Productivity Measurement With Adjustments for Variations in Capacity Utilization, and Other Forms of Temporary Equilibrium," Journal of Econometrics, 33, 1986, 7-29.

Berndt, Ernest R. (1980), "Energy Price Increases and the Productivity Slowdown in U.S. Manufacturing," The Decline in Productivity Growth, Federal Reserve Bank of Boston, Conference Series No. 22, Boston, 60-89.

Cartwright, David W., "Improved Deflation of Purchases of Computers," Survey of Current Business, vol. 66, no. 3, March 1986, 7-9.

Cas, Alexandra and Thomas K. Rymes, On Concepts of Multifactor Productivity, Cambridge University Press, New York, 1991.

Cass, David, "Optimum Growth in an Aggregative Model of Capital Accumulation," Review of Economic Studies, 32, July 1965, 233-240.

Caves, Douglas W., Laurits R. Christensen, and W. Erwin Diewert, "Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers" Economic Journal, 92, March 1982a, 73-86.

Caves, Douglas W., Laurits R. Christensen, and W. Erwin Diewert, "The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity," Econometrica, 50, 6, November 1982b, 1393-1414.

Christensen, Laurits R., Diane Cummings, and Dale W. Jorgenson, "Relative Productivity Levels, 1947-1973: An International Comparison," European Economic Review, 16, 1981, 61-94.

Christensen, Laurits R. and Dale W. Jorgenson, "The Measurement of U.S. Real Capital Input, 1929-1967," Review of Income and Wealth, 15, December 1969, 293-320.

Christensen, Laurits R. and Dale W. Jorgenson, "U.S. Real Product and Real Factor Input, 1929-1969," Review of Income and Wealth, 16, March 1970, 19-50.

Christensen, Laurits R., Dale W. Jorgenson and Lawrence J. Lau, "Transcendental Logarithmic Production Frontiers," Review Economics and Statistics, 55, February 1973, 28-45.

Cobb, Clifford, Ted Halstead, and Jonathan Rowe, "If the GDP is Up, Why is America Down?, The Atlantic Monthly, 276, 4, October 1995, 59-78.

Cole, Rosanne, Chen, Y.C., Barquin-Stolleman, J.A., Dullberger, E., Helvacian, N., and Hodge, J.H., "Quality-Adjusted Price Indexes for Computer Processors and Selected Peripheral Equipment," Survey of Current Business, 66, January 1986, 41-50.

Collins, Susan M., and Barry P. Bosworth, "Economic Growth in East Asia: Accumulation versus Assimilation," Brookings Papers on Economic Activity, 2, 1996, 135-191.

Copeland, Morris A., "Concepts of National Income," Studies in Income and Wealth, volume 1, New York, National Bureau of Economic Research, 1937, 3-63.

Copeland, Morris A., and E. M. Martin, "The Correction of Wealth and Income Estimates for Price Changes," volume 2, New York, National Bureau of Economic Research, 1938, 85-135 (including discussion by Milton Friedman).

Davis, Steven J., and John C. Haltiwanger, "Wage Dispersion Between and Within U.S. Manufacturing Plants, 1963-86," Brookings Papers on Economic Activity, 1, 1991, 115-120.

Denison, Edward F., The Sources of Economic Growth in the United States and the Alternatives Before Us, New York: Committee for Economic Development, 1962.

Denison, Edward F., "The Unimportance of the Embodiment Question," American Economic Review, 79, No. 5, March 1964, 90-94.

Denison, Edward F., "Some Major Issues in Productivity Analysis: an Examination of the Estimates by Jorgenson and Griliches," Survey of Current Business, 49 (5, Part II), 1972, 1-27.

Denison, Edward F., "Explanations of Declining Productivity Growth," Survey of Current Business, 59, August 1979, 1-24.

Denison, Edward F., Estimates of Productivity Change by Industry: An Evaluation and an Alternative, Washington, D.C.: The Brookings Institution, 1989.

Denny, Michael, Melvyn Fuss, and Leonard Waverman, "The Measurement and Interpretation of Total Factor Productivity in Regulated Industries, with an Application to Canadian Telecommunications," in T. Cowing and R. Stevenson, eds., Productivity Measurement in Regulated Industries, Academic Press, New York, 1981, 179-218.

Diewert, W. Erwin, "Aggregation Problems in the Measurement of Capital," in Dan Usher, ed., The Measurement of Capital, National Bureau of Economic Research, Studies in Income and Wealth, volume 45, University of Chicago Press, Chicago, 1980, 433-528.

Diewert, W. Erwin, "Exact and Superlative Index Numbers," Journal of Econometrics, 4, 1976, 115-145.

Diewert, W. Erwin, and Kevin J. Fox, "Can Measurement Error Explain the Productivity Paradox," Canadian Journal of Economics, 32, April 1999, 251-280.

Domar, Evsey D., "On the Measurement of Technical Change," Economic Journal, 71, 1961, 710-729.

Domar, Evsey D., "Total Factor Productivity and the Quality of Capital," Journal of Political Economy, 71, December 1963, 586-588.

Dowrick Steven and Duc-Tho Nguyen, "OECD Comparative Economic Growth 1950-85: Catch-up and Convergence," American Economic Review, 79, No. 5, December 1989, 1010-1030.

Easterly, William, "The Mystery of Growth: Shocks,Policies, and Surprises in Old and New Theories of Economic Growth," Singapore Economic Review, 40, 1, 1995, 3-23.

Färe, Rolf, Shawna Grosskopf, Mary Norris, and Zhongyang Zhang, "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries," American Economic Review, 84, 1, March 1994, 66-83.

Fisher, Franklin, "Embodied Technical Change and the Existence of an Aggregate Capital Stock," Review of Economic Studies, 32, 1965, 326-388.

Gallman, Robert E., "Investment Flows and Capital Stocks: U.S. Experience in the 19th Century," in Peter Kilby, ed., Quantity and Quiddity: Essays in U.S. Economic History in Honor of Stanley Lebergott, Middletown, Connecticut, Wesleyan University Press, 1987, 214-254.

Goldsmith, Raymond W., The National Wealth of the United States in the Postwar Period, Princeton: Princeton University Press, 1962.

Gordon, Robert J., The Measurement of Durable Goods Prices, Chicago: University of Chicago Press for the National Bureau of Economic Research, 1990.

Gore, Albert Jr., Earth in the Balance: Ecology and the Human Spirit, Houghton Mifflin, Boston, 1992.

Greenwood, Jeremy, Zvi Hercowitz, and Per Krusell, "Long-Run Implications of Investment-Specific Technical Change," American Economic Review, 87, 3, June 1997, 342-362.

Griliches, Zvi, "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change," in, The Price Statistics of the Federal Government, New York: National Bureau of Economic Research, General Series, 73, 1961, 137-196.

Griliches, Zvi, "Notes on the Measurement of Price and Quality Changes," in Models of Income Determination, Studies in Income and Wealth, Vol. 28, Princeton University Press for NBER, Princeton, N.J., 1964, 381-404.

Griliches, Zvi, "Research Expenditures and Growth Accounting," in B.R. Williams, ed., Science and Technology in Economic Growth, The Price Statistics of the Federal Government, MacMillan Press, 1973.

Griliches, Zvi, "R&D and the Productivity Slowdown," American Economic Review, 70, No. 2, May 1980, 343-348.

Griliches, Zvi, "The Search for R&D Spillovers," Scandinavian Journal of Economic, 94, 1992, 29-47.

Griliches, Zvi, "Productivity, R&D, and the Data Constraint," American Economic Review, 84, 1, March 1994, 1-23.

Griliches, Zvi, "The Discovery of the Residual: A Historical Note," Journal of Economic Literature, 34, September 1996, 1324-1330.

Griliches, Zvi, and Jacques Mairesse, "Productivity and R&D Growth at the Firm Level," in Zvi Griliches, ed., R&D, Patents, and Productivity, Chicago: Chicago University Press, 1984, 339-374.

Grossman, Gene M. and Elhanan Helpman, Innovation and Growth in the Global Economy, MIT Press, Cambridge Massachusetts, 1991.

Gullickson, William, and Michael J. Harper, "Possible Measurement Bias in Aggregate Productivity Growth," Monthly Labor Review, forthcoming.

Hall, Bronwyn H., "New Evidence on the Impacts of R&D," Brookings Papers on Economic Activity, Microeconomics, 1993.

Hall, Robert E., "Technical Change and Capital From the Point of View of the Dual," Review of Economic Studies, 35, 1968, 34-46.

Hall, Robert E., "The Relation Between Price and Marginal Cost in U.S. Industry," Journal of Political Economy, 96, 1988, 921-947.

Hall, Robert E. and Dale W. Jorgenson, "Tax Policy and Investment Behavior," American Economic Review, 57, 1967, 391-414.

Haltiwanger, John, "Measuring and Analyzing Aggregate Fluctuations: The Importance of Building from Microeconomic Foundations," Federal reserve Bank of St. Louis Economic Review, January/February, 1997.

Hicks, John, Value and Capital, Oxford University Press, London, 1946.

Hulten, Charles R., "Divisia Index Numbers," Econometrica, 41, 1973, 1017-1025.

Hulten, Charles R., "Technical Change and the Reproducibility of Capital," American Economic Review, 65, 5, December 1975, 956-965.

Hulten, Charles R., "Growth Accounting with Intermediate Inputs," Review of Economic Studies, 45, October 1978, 511-518

Hulten, Charles R., "On the 'Importance' of Productivity Change," American Economic Review, 69, 1979, 126-136.

Hulten, Charles R., "Productivity Change, Capacity Utilization and the Source of Efficiency Growth," Journal of Econometrics, 33, 1986, 31-50.

Hulten, Charles R, "The Measurement of Capital," in Ernst R. Berndt and Jack E. Triplett, eds., Fifty Years of Economic Measurement, Studies in Income and Wealth, Volume 54, Chicago: Chicago University Press for the National Bureau of Economic Research, 1990, 119-152.

Hulten, Charles R., "Accounting for the Wealth of Nations: The Net versus Gross Output Controversy and its Ramifications," Scandinavian Journal of Economics, 94, 1992a (supplement), S9-S24.

Hulten, Charles R., "Growth Accounting when Technical Change is Embodied in Capital," American Economic Review, Vol 82, No. 4, September 1992b, 964-980.

Hulten, Charles R., "Comment on 'Do Real Output and Real Wage Measures Capture Reality? The History of Lighting Suggests Not'," Timothy Bresnahan and Robert J. Gordon, eds., The Economics of New Goods, Studies in Income and Wealth, Volume 58, Chicago: University of Chicago Press for the National Bureau of Economic Research, 1997, 66-70.

Hulten, Charles R., "Quality Change in Capital Goods and its Impact on Economic Growth," National Bureau of Economic Research Working Paper 5569, May 1996.

Hulten, Charles R., and Sylaja Srinivasan, "Indian Manufacturing Industry: Elephant or Tiger," University of Maryland, October 1999.

Hulten, Charles R. and Frank C. Wykoff, "The Estimation of Economic Depreciation Using Vintage Asset Prices," Journal of Econometrics, 15, 1981, 367-396.

Johansen, Leif, "Substitution versus Fixed Production Coefficients in the Theory of Economic Growth: A Synthesis," Econometrica, 27, April 1959, 157-176.

Jones, Charles I., "Times Series Tests of Growth Models," Quarterly Journal of Economics, May 1995, 495-525.

Jones, Charles I., "R&D-Based Models of Economic Growth," Journal of Political Economy, 103, August 1995, 759-784.

Jorgenson, Dale W., "Capital Theory and Investment Behavior," American Economic Review, 53, 2, May 1963, 247-259

Jorgenson, Dale W., "The Embodiment Hypothesis," Journal of Political Economy, 74, February 1966, 1-17.

Jorgenson, Dale W., Frank M. Gollop, and Barbara M. Fraumeni, Productivity and U.S. Economic Growth, Cambridge, MA, The Harvard University Press, 1987.

Jorgenson, Dale W. and Barbara M. Fraumeni, "The Accumulation of Human and Nonhuman Capital, 1948-84," in Robert E. Lipsey and Helen Stone Tice, eds., The Measurement of Saving, Investment, and Wealth, Studies in Income and Wealth, Volume 52, Chicago University Press for the National Bureau of Economic Research, Chicago, 1990, 119-152.

Jorgenson, Dale W. and Zvi Griliches, "The Explanation of Productivity Change," Review of Economic Studies, 34, July 1967, 349-83.

Jorgenson, Dale W. and Zvi Griliches, "Issues in Growth Accounting: A Reply to Edward F. Denison," Survey of Current Business, 52, 65-94, 1972.

Jorgenson, Dale W. and Mieko Nishimizu, "U.S. and Japanese Economic Growth, 1952-1974: An International Comparison," Economic Journal, 88, December, 1978, 707-726.

Kendrick, John, Productivity Trends in the United States, New York: National Bureau of Economic Research, 1961.

Kim, Jong-Il, and Lawrence J. Lau, "The Sources of Economic Growth of the East Asian Newly Industrialized Countries," Journal of Japanese and International Economies, 8, 1994, 235-27.

Koopmans, T.C., "On the Concept of Optimal Economic Growth," Pacifica Academia Scientiarus, Rome, 1965, 276-279.

Krugman, Paul, "The Myth of Asia's Miracle," Foreign Affairs, 73, 6, November/December, 1994, 62-77.

Lucas, Robert E. Jr., "On the Mechanics of Economic Development," Journal of Monetary Economics, 22, 1988, 3-42.

Maddison, Angus, "Growth and Slowdown in Advanced Capitalist Economies: Techniques and Quantitative Assessment," Journal of Economic Literature, 25, 2, June 1987, 649-698.

Nadiri, M. Ishaq, "Some Approaches to the Theory and Measurement of Total Factor Productivity: A Survey," Journal of Economic Literature, 8, December, 1970, 1137-1177.

Nadiri, M. Ishaq, and Seongjun Kim, "R&D, Production Structure, and Productivity Growth: A Comparison of US, Japan, and Korean Manufacturing Sectors," RR #96-11, C.V. Starr Center for Applied Economics, New York University, March 1996.

Nadiri, M. Ishaq, Innovations and Technological Spillovers, Working Paper No. 4423, National Bureau of Economic Research, Cambridge, Massachusetts, August 1993.

Nelson, Richard R., "Aggregate Production Functions and Medium-Range Growth Projections, American Economic Review, 54, 1964, 575-606.

Nordhaus, William D., "Do Real Output and Real Wage Measures Capture Reality? The History of Lighting Suggests Not," Timothy Bresnahan and Robert J. Gordon, eds., The Economics of New Goods, Studies in Income and Wealth, Volume 58, Chicago: University of Chicago Press for the National Bureau of Economic Research, 1997, 29-66.

Nordhaus, William D., and Edward C. Kokkelenberg, eds., Nature's Numbers, Expanding the National Economic Accounts to Include the Environment, National Research Council, National Academy Press, Washington, D.C., 1999.

Prucha, Ingmar R., and M. Ishaq Nadiri, "Endogenous Capital Utilization and Productivity Measurement in Dynamic Factor Demand Models: Theory and an Application to the U.S. Electrical Machinery Industry," Journal of Econometrics, 15, 1981, 367-396.

Repeto, Robert, Dale Rothman, Paul Faeth, and Ducan Austin, Has Environmental Protection Really Reduced Productivity Growth? World Resources Institute, Washington, D.C., October 1996.

Richter, Marcel K., "Invariance Axioms and Economic Indexes," Econometrica, 34, 1966, 739-755.

Rodrik, Dani, "TFPG Controversies, Institutions and Economic Performance in East Asia," in Y. Hayami and N. Aoki, EDS., The Institutional Foundations of Economic Development in East Asia, London, MacMillan, 1998.

Romer, Paul M., "Increasing Returns and Long-Run Growth," Journal of Political Economy, 94, No. 5, 1986, 1002-1037.

Rosen, Sherwin, "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," Journal of Political Economy, 82, January-February 1974, 34-55.

Rymes, Thomas K., On Concepts of Capital and Technical Change, Cambridge, Cambridge University Press, 1971.

Salter, W.E.G., Productivity and Technical Change, Cambridge: Cambridge University Press, 1960.

Solow, Robert M.,"A Contribution to the Theory of Economic Growth," Quarterly Journal of Economics, 70, February 1956, 65-94.

Solow, Robert M., "Technical Change and the Aggregate Production Function," Review of Economics and Statistics, 39, August 1957, 312-320.

Solow, Robert M., "Investment and Technical Progress," in K. Arrow, S. Karlin and P. Suppes, eds., Mathematical Methods in the Social Sciences, Stanford University Press, 1960.

Solow, Robert M., Book Review, New York Times, July 12, 1987, 36.

Stigler, George J., Trends in Output and Employment, New York, National Bureau of Economic Research, 1947.

Swan, T.W. "Economic Growth and Capital Accumulation," The Economic Record, XXXII, 1956, 334-361.

Tinbergen, Jan, "Zur Theorie der Langfirstigen Wirtschaftsentwiicklung," Weltwirst. Archiv., 1, Amsterdam, North-Holland Publishing Company, 1942, 511-549.

Tornqvist, L., "The Bank of Finland's Consumption Price Index," Bank of Finland Monthly Bulletin, 10, 1936, 1-8.

Triplett, Jack E., "Concepts of Quality in Input and Output Price Measures: A Resolution of the User Value-Resource Cost Debate," in Murray F. Foss, ed., The U.S. National Income and Product Accounts: Selected Topics, Studies in Income and Wealth, Volume 47, Chicago: University of Chicago Press for the National Bureau of Economic Research, 1983, 296-311.

Triplett, Jack E., "Hedonic Functions and Hedonic Indexes," in John Eatwell, Murray Milgate and Peter Newman, eds., The New Palgrave Dictionary of Economics, Vol. 2, New York: The Macmillan Press Limited, 1987, 630-634.

Triplett, Jack E., "Depreciation in Production Analysis and Economic Accounts," Economic Inquiry, January 1996, 93-115.

U.S. Department of Labor, Bureau of Labor Statistics, Trends in Multifactor Productivity, 1948-81, Bulletin 2178, USGPO, Washington D.C., September 1983.

U.S. Department of Commerce, Bureau of The Census, Historical Statistics of the United States, Colonial Times to 1970, USGPO, Washington D.C., 1975.

Weitzman, Martin L., "On the Welfare Significance of National Product in a Dynamic Economy," The Quarterly Journal of Economics, 90, 1976, 156-162.

Wolff, Edward, N., "The Productivity Slowdown: The Culprit At Last? Follow-up on Hulten and Wolff," American Economic Review, 86, 5, December 1996, 1239-1252.

Young, Alwyn, "A Tale of Two Cities: Factor Accumulation and Technical Change in Hong Kong and Singapore," in Blanchard, O.J., and S. Fisher (eds.), NBER Macroeconomics Annual 1992, MIT Press, Cambridge MA, 13-53.

Young, Alwyn, "The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Experience," The Quarterly Journal of Economics, August 1995, 641-680.

Zucker, Seymore, "Low Prices. Low Rates. It's Time to Relax.", Business Week, September 15, 1997, 37.

TABLE 1

HISTORICAL GROWTH RATES OF OUTPUT PER PERSON
AND TOTAL FACTOR PRODUCTIVITY IN THE U.S.
(BY DECADE)

DECADE	REAL GNP/GDP PER CAPITA	TFP	CONTRIBUTION OF TFP
1779-1789	-0.002	N/A	
1789-1799	-0.008	N/A	
1799-1809	0.007	0.006	73.5%
1809-1819	-0.009	0.006	64.4%
1819-1829	0.008	0.006	69.7%
1829-1839	0.012	0.006	44.0%
1839-1849	0.018	0.007	38.4%
1849-1859	0.016	0.007	45.1%
1859-1869	0.004	0.007	161.7%
1869-1879	0.023	0.007	30.7%
1879-1889	0.017	0.007	42.7%
1889-1899	0.023	0.003	12.6%
1899-1909	0.018	0.002	13.5%
1909-1919	0.019	0.003	16.3%
1919-1929	0.024	0.002	7.7%
1929-1939	0.016	0.003	16.6%
1939-1949	0.026	0.003	9.6%
1949-1959	0.034	0.002	6.2%
1959-1969	0.027	0.003	12.0%
1969-1979	0.023	N/A	
1979-1989	0.017	N/A	
1989-1997	0.009	N/A	
1799-1979	0.018	0.005	26.0%
PRIVATE BUSINESS ECONOMY ONLY			
1948-1973	0.033	0.021	64%
1973-1979	0.013	0.006	46%
1979-1990	0.012	0.002	17%
1990-1996	0.011	0.003	27%
1948-1996	0.023	0.012	52%

Sources: Gallman (1987), the <u>Historical Statistics of the United States, Colonial Times to 1970</u>, and the 1998 <u>Economic Report of the President</u>. Data in lower panel are from the Bureau of Labor Statistics.

TABLE 2

SOURCES OF GROWTH IN U.S. PRIVATE BUSINESS SECTOR
(SELECTED INTERVALS)

Years	Real Output	Labor Input	Capital Services	TFP
1948-1996	3.4	1.4	3.7	1.2
1948-1973	4.0	1.0	3.8	2.1
1973-1996	2.7	1.9	3.5	0.3
1973-1979	3.1	1.8	4.1	0.6
1979-1990	2.7	2.0	3.8	0.2
1990-1996	2.4	1.9	2.5	0.3

Source: Bureau of Labor Statistics Multifactor Productivity Program

Figure 1

Real GNP/GDP Per Capita in U.S.