

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: NBER Macroeconomics Annual 1987, Volume 2

Volume Author/Editor: Stanley Fischer, editor

Volume Publisher: The MIT Press

Volume ISBN: 0-262-56040-0

Volume URL: <http://www.nber.org/books/fisc87-1>

Publication Date: 1987

Chapter Title: Crazy Explanations for the Productivity Slowdown

Chapter Author: Paul M. Romer

Chapter URL: <http://www.nber.org/chapters/c11101>

Chapter pages in book: (p. 163 - 210)

Paul M. Romer

UNIVERSITY OF ROCHESTER

Crazy Explanations for the Productivity Slowdown

1. Introduction

In 1958 Wolfgang Pauli presented a paper in which he and his coauthor, Werner Heisenberg, tried to solve all of the outstanding problems in elementary particle physics with a single equation. This article tries to do the same for the productivity slowdown in the United States—explain everything in terms of a single growth accounting equation. Given the amount of effort that has been devoted to growth accounting and the productivity puzzle, it would be surprising if this effort were a complete success. One suspects that any area that has received this much attention will be subject to a conservation law, the conservation of puzzles.

The analogy with particle physics in the first half of this century strikes me as apt because there appears to be a wide range of fundamental problems confronting the neoclassical theory of growth that forms the foundation for growth accounting and productivity analysis.¹ I take the key features of that theory to be perfect competition, a constant-returns-to-scale production function, and a specification of technologi-

I would like to thank without implicating, Mark Bills, Eliana Cardoso, Stanley Engerman, Robert Gordon, Robert E. Lucas Jr., Walter Oi, Edward Prescott, Sergio Rebelo, Ken Rogoff, Christina Romer, Larry Rosenblum, Lawrence Summers; the discussants, Martin Baily and Ben Bernanke; and the participants in the discussion for many useful comments, suggestions, and criticisms, not all of which could be reflected in the revised version of the paper.

1. Examples of these problems are the failure of low-income countries to catch up with high-income countries, the fact that returns to capital do not seem to vary with capital per worker by as much as theory would suggest, and the persistent attempts by labor to migrate toward capital despite the relatively unrestricted ability of capital to move toward labor. See Lucas (1985), Romer (1986a), and De Long (1987) for a discussion of these issues.

cal change as an exogenous function of time or as an ad hoc function of variables that can be analyzed in isolation from the basic factor inputs. Models in which the direction of technological change or the overall rate of change is responsive to economic incentives have been proposed,² but they have had little impact on empirical practice, in part because the underlying theory is typically incomplete.

The comment offered by Neils Bohr after Pauli finished his presentation also strikes me as apt. "Herr Professor," said Bohr, "we are all agreed that your theory is crazy. However, we are not convinced that it is crazy enough to be right."³ Consequently, I will try to offer suggestions that are not just crazy, but are radically so. For example, I will argue that we should drop the notion of technological change altogether and work with a production function that can be described as a stationary function of measurable inputs. To make this description consistent with the data, it will be necessary to drop the assumptions required for perfect competition, especially the assumption that there are no external effects. This implies that one cannot simply equate prices with marginal products and account for growth by weighting growth rates of different inputs by factor shares in total income. I will allow for the possibility that there are aggregate increasing returns, and I will even argue that an important part of technological change may act to lower the marginal productivity of labor.

In part, I take this route because all the sensible suggestions concerning the productivity slowdown have already been explored. I am in substantial agreement with many of these suggestions, but the emphasis here is on the basic theoretical foundation for growth accounting, not on the details of possible refinements. As a result, I will use a framework that, except for the absence of constant returns to scale, perfect competition, and technological change, is very similar to that used by Solow (1957). I will not take into account the role of inputs other than capital or labor. I will look only at a value-added measure of output, for the most part at a fairly high degree of aggregation.⁴ I will try to use the most sensible measures of capital and labor inputs available from other sources, but I will not make my own adjustments to the existing measures to take

2. For models of the first type see Fellner (1961), Kennedy (1964), Samuelson (1965), Drandakis and Phelps (1966); for the second type, see Phelps (1966), von Weizsacker (1966), Shell (1967).

3. This lecture is described in a recent article by Jeremy Bernstein (*New Yorker*, Feb. 2, 1987). The quotation given here was described to me by Leo Nedelski, a physicist at the University of Chicago. Bohr was right.

4. See Bruno (1984) and Baily (1986) for a discussion of the role of materials inputs and net versus gross output measures. See Jorgenson, Gollop, and Fraumeni (1986) for detailed estimates using gross output at the industry level.

account of changes in quality of labor,⁵ nor will I consider the effects of alternative methods of aggregation and depreciation for the measured stock of capital.⁶ I will not offer any new evidence at the level of individual industries, nor consider the possibility that depletion of the stock of technological innovations explains the slowdown in particular industries.⁷ I will not look at any direct measures of research, development, or inventive activity.⁸ I will not offer anything beyond the usual handwringing about mismeasured quality change in the division of nominal income growth into real growth and inflation.⁹ Much careful work on these issues has been and continues to be done, and nothing suggested here makes any of them less important. Because there is no consensus about the appropriate explanation for recessions, I will follow the older tradition in the literature on productivity and maintain a theoretical dichotomy between growth and business cycles.¹⁰

Section 2 describes three distinct theoretical models that allow for external effects and for the possibility of a departure from constant returns to scale. The intention in this section is not to produce a fully formed model with parameters to be estimated and restrictions to be tested. Rather, the intention is to suggest features of the data that might be worth examining and to induce a predisposition to let the data speak for themselves before we force the elasticity of output with respect to capital or labor to be equal to that factor's share in total income. Section 3, which summarizes the evidence, therefore amounts to a rather loosely structured form of data analysis. It considers not just data on the productivity slowdown per se, but also more general evidence on the plausibility of the alternative theories offered here and on the elasticity of output with respect to capital and labor inputs.

Before asking the reader to invest in the lengthy arguments that follow, it is reasonable to offer a preview of the conclusions they lead to. First, it appears that the conventional growth accounting analysis substantially underestimates the role of capital accumulation in growth. The correct weight on the rate of growth of capital in a growth accounting exercise

5. See Darby (1984) or Gollop and Jorgenson (1980) for estimates that rely on measures of labor adjusted for quality change.
6. See Norsworthy (1984) for a detailed comparison of how capital is measured in the growth accounting approaches of Denison, Kendrick, and Jorgenson.
7. For examples of research along these lines, see Gordon (1983), and Baily and Chakrabarti (1985).
8. See Griliches (1980, 1986), Sveikauskas (1986).
9. For examples of research that does treat this issue seriously, see Lichtenberg and Griliches (1986) and Gordon (1986b).
10. Recent work that does not maintain this dichotomy includes Hall (1986), Prescott (1986), and King and Rebelo (1986).

may be closer to 1 than to 0.25. The true elasticity of output with respect to changes in capital may be greater than the share of capital in total income because of positive externalities associated with investment.

This view is consistent with the long-run growth of output and productivity and can explain growth without introducing exogenous technological change. However, it completely resolves neither the postwar acceleration of productivity and output growth in the United States nor the subsequent slowdown in growth. One way to shift some of the blame for the decade-to-decade variations may be to implicate business cycle fluctuations. It does appear that they are more important and harder to remove from the data than is conventionally recognized. One revealing piece of evidence in this regard is the fact that the labor productivity slowdown for manufacturing in the United States, which seemed so obvious and persistent during the late seventies and early eighties, has disappeared over the course of the current expansion.

Even combined with an allowance for cyclical fluctuations, emphasis on the elasticity of output with respect to capital cannot by itself explain the behavior of labor productivity in the United States, especially in non-manufacturing over the last thirty years. There, the data seem to be telling us that output responds much less to increases in the amount of labor supplied than a simple model like $Y = K^\alpha L^{1-\alpha}$ would suggest, whatever value we assign to α . Long-run data suggest that the exponent on labor may be substantially smaller than its share in income, possibly on the order of 0.1 or 0.2. If estimates in this range are correct, a *ceteris paribus* increase of 1 percent per year in the rate of growth of labor will cause a fall of 0.8 percent or 0.9 percent per year in the rate of growth of labor productivity.

One interpretation of such a large difference between the elasticity of output with respect to labor and labor's share in income is that there is a negative externality associated with labor. This could arise if there is a form of innovation that economizes on labor, if investment in this kind of innovation is sensitive to movements in wages, and if this innovation has positive external effects because of spillovers of knowledge. In this case, an increase in the rate of growth of the labor force, with the implied decrease in the rate of growth of wages, could cause a decrease in innovation, and hence a decrease in knowledge spillovers from innovation. The net effect that an increase in labor supply has on output would then be the combination of the positive direct effect of more workers and the negative indirect effect of less innovation.

The suggestion that this kind of effect could be present is not new.¹¹

11. Suggestions along these lines are contained for example in Hicks (1963).

This kind of interaction between wages and innovation has been invoked repeatedly in the comparative analysis of productivity growth in the United States and Britain during the nineteenth and early twentieth centuries.¹² What seems at variance with reasonable prior beliefs is the suggestion that the magnitude of this effect could be so large. As a result, the exercise undertaken here does not fully resolve the productivity puzzle, but rather converts it into a different, possibly more suggestive, puzzle. Reconciling this explanation with these apparently reasonable prior beliefs, or finding an alternative explanation for the low apparent elasticity of output with respect to labor is the new puzzle suggested here.

2. Theory

2.1. THE SOLOW MODEL

Because it is familiar, the implications of the basic Solow model will be noted only briefly. The central equation is

$$Y(t) = \Gamma(t)F(K(t),L(t)), \quad (1)$$

where F is assumed to exhibit constant returns to scale and $\Gamma(t)$ represents exogenous technological change. Under the assumption of competition, it follows immediately that percentage growth rates, denoted here by variables with "^^"s, can be written

$$\hat{Y} = \hat{\Gamma} + \alpha \hat{K} + (1 - \alpha) \hat{L}, \quad (2)$$

with α defined to be the share of capital in total income.

In applications, this model is used in two quite different ways. For a single developed country, analysis often focuses on the behavior of steady-state growth, ignoring any transitional dynamics. Under the simplifying assumptions that the function F can be taken to be Cobb-Douglas and that technological change is exponential— $\Gamma(t) = e^{\gamma t}$ —steady-state growth at a rate $\hat{Y} = \hat{K} = \hat{L} + \gamma/(1 - \alpha)$ will occur if a constant fraction of income is saved or if preferences take the usual additively separable form. According to this view, the rate of growth of output per worker, $\hat{Y} - \hat{\Gamma}$, is determined entirely by the rate of technological change. A ceteris paribus increase in \hat{L} will be matched one for one by an increase in \hat{Y} .

The empirical applicability of these steady-state results is not clear.

12. The early references in this area are Rothbart (1946) and Habakkuk (1962).

Sato (1963) and Sato (1966) show that the half-life of the transitional dynamics in a simple growth model can be quite long, on the order of many decades. Also, evidence presented in the next section on changes in output in the United States in response to changes in labor supply shows that the steady-state results may not offer a reasonable description of the data when growth rates are measured over periods as long as twenty years.

Alternatively, the model is interpreted as a description of the behavior of a cross-section of economies under the assumption that they are not all on steady-state growth paths. Countries starting from a low level of the capital-labor ratio are assumed to accumulate capital more rapidly and catch up with countries that start from a higher initial position. For example, countries like Germany and Japan that suffered large losses during World War II, or developing countries like Korea, are thought to have grown faster than the United States in the 1950s and 1960s because of more rapid capital accumulation as they approached the capital-labor ratio in the United States. Under the assumption that the rate of technological change $\hat{\Gamma}$ is the same in each country, the difference in the rate of growth between countries 1 and 2 can be expressed as

$$\hat{Y}_1 - \hat{Y}_2 = \alpha(\hat{K}_1 - \hat{K}_2) + (1 - \alpha)(\hat{L}_1 - \hat{L}_2). \quad (3)$$

One of the implications of this model, noted by Robert Solow (Solow 1959) is that a value of α near $\frac{1}{4}$ or even $\frac{1}{3}$ leaves little scope for increasing the rate of growth of output by increasing capital accumulation. Increasing the investment-output ratio I/Y by Δ leads to an increase in the rate of growth of output equal to $\alpha\Delta(Y/K)$. Thus for $Y/K = \frac{1}{3}$, an increase in the share of output devoted to investment of 10 percentage points leads to an increase in the growth rate of output of roughly 1 percent per year. For this reason, and also because of the implausible suggestion that exogenous technological change would cause output to increase in the absence of any increment in the stocks of labor and capital, Solow suggested that technological change must be embodied in new investment. In effect, this is equivalent to assuming that technological change arises entirely through unmeasured quality change in capital inputs.

To see how this changes the model, let output be written as $Y = F(J, L)$, where J is the true aggregate measure of capital in efficiency units. Let $I(t) = Y(t) - C(t)$ denote forgone output devoted to investment in new capital goods. Assuming exponential depreciation at the rate δ , the equation for the accumulation of capital is $\dot{J}(t) = A(t)I(t) - \delta J(t)$. Exogenous technological change is captured entirely in the term

$A(t)$. If the measured stock of capital is simply a direct aggregate of forgone consumption, $\hat{K}(t) = I(t) - \delta K(t)$, it measures capital at cost and makes no correction for the fact that one unit of forgone consumption today makes a better capital good than one unit yesterday; that is, it takes no account of the quality change associated with increases in $A(t)$.

In this model, α , the elasticity of output with respect to changes in true capital J , is correctly measured by the share of income devoted to capital. As shown in Solow (1956), all that is required for this result is competition and constant returns.¹³ The difficulty introduced by embodied technological change lies in the measurement of \hat{J}/J . The growth rate of capital as conventionally measured, $\hat{K}/K = I/K - \delta$, underestimates the effect of an increase in I because the measure K counts older, less productive capital goods equally with more recent goods.

To see this in a simple example, assume that $A(t)$ has been growing forever at an exponential rate a , that $I(t)$ has been growing at the rate g for all t prior to time 0, and that $I(0)$ is normalized to 1. At time 0, $K(0) = 1/(g + \delta)$, $J(0) = A(0)/(g + a + \delta)$. Then if I is chosen freely at time s just after 0, $\hat{J} = A(s)I - \delta J$ and $\hat{K} = I - \delta K$. Therefore, $\hat{J}/J = (g + a + \delta)I(s) - \delta$, but $\hat{K}/K = (g + \delta)I(s) - \delta$. A change in I therefore leads to a change in \hat{J}/J that differs from the change in \hat{K}/K by the factor

$$\lambda = \frac{A(0)K(0)}{J(0)} = \frac{g + a + \delta}{g + \delta}. \quad (4)$$

An estimate of the elasticity of output with respect to capital based on a regression of $\hat{Y} - \hat{L}$ on \hat{K} instead of \hat{J} would be too low by the factor λ^{-1} . This same factor arises in the calculation of the effect that a change in the share of GNP devoted to investment has on growth. The growth rate of output can be written $\hat{Y} = \hat{F} - \alpha\delta + \alpha(I/Y)(AK/J)(Y/K) + (1 - \alpha)\hat{L}$. Then using the definition of λ , a change Δ in I/Y leads to a change in \hat{Y} of $\alpha\Delta(Y/K)\lambda$.

The magnitude of this effect depends on the values for the parameters g , δ , and a . Using direct estimates of characteristics of capital goods, Gordon (1986b) estimates that existing deflators for capital goods may underestimate quality change by as much as 3 percent per year. Lichtenberg and Griliches (1986) use a comparison of different producer price indices to conclude that the mismeasurement is on the order of 1.5 percent per year. Combining this range of estimates for a with an estimate of

13. These assumptions are not sufficient to imply that α is a constant, but since it is observable, this causes no problem for the implementation of growth accounting.

the growth rate g for capital of 2 percent per year and an estimate of depreciation of 4 percent per year, the factor λ lies in the range 1.25 to 1.50. Thus, even if the true elasticity α is as large as $\alpha = .33$, the estimate of α generated by regressing the growth rate of output \hat{Y} on growth rates of \hat{K} and \hat{L} should be no larger than 0.5. Similarly, adding 10 percentage points to I/Y will lead to at most a 1.5 percent increase in \hat{Y} , as opposed to the previous estimate of 1.0 percent. If the rate of growth of capital or the rate of depreciation is larger, the factor λ will be even smaller.

2.2. GROWTH WITH INCREASING RETURNS AND KNOWLEDGE SPILLOVERS

The only plausible way to get something resembling a direct external effect associated with investment decisions by firms is to allow for the possibility that knowledge is an intangible capital input in production. Thus, $F(K, L, e)\Omega(E)$ could represent the production possibilities available to a firm with capital K , labor input L and a private stock of knowledge e when the aggregate stock of public knowledge is E . For example, e could represent process knowledge that is kept secret by a firm, whereas E could represent the publicly observable outcome of product design efforts by all firms in the economy. An individual firm that introduces a new good does research to produce new knowledge about designs and about processes for producing it. The implemented design knowledge becomes part of the aggregate pool E that any firm can exploit.

In the simplest fixed proportions case, E is assumed to increase one-for-one with e , and e is in turn assumed to move one-for-one with K .¹⁴ With identical firms in an economy, production can then be written $Y = F(K, L)\Omega(K)$. If F is Cobb-Douglas and Ω is the power function $\Omega(K) = K^\theta$, aggregate output takes the form

$$F(K, L)\Omega(K) = K^\alpha L^{1-\alpha} K^\theta. \quad (5)$$

If the stock of labor is fixed, if $\alpha + \theta$ is less than one, and if capital is accumulated one-for-one from forgone consumption, then diminishing returns to investment in capital will eventually force growth to stop just as it does in the classical case. One way to generate unceasing growth is to allow for exogenous growth in the labor force. As K grows in proportion with L , output per capita will increase over time because of the increasing returns to scale. This is essentially the mechanism proposed by

14. The first assumption is essentially a matter of convenience. Qualitatively, all that matters is that public knowledge increases with private knowledge. Section 3.1 describes evidence presented in Schmookler (1966) on the comovements of patents and investment, which suggests that the second assumption may not be unreasonable.

Arrow (1962). The alternative proposed in my earlier paper (Romer 1986a) is to allow for the possibility that $\alpha + \theta$ is greater than or equal to 1, so that the per capita capital stock can increase without bound. A simple intermediate case is where $\alpha + \theta$ equals 1, so aggregate production is $F(K,L) = KL^{1-\alpha}$. With a fixed labor force, the model will behave much like a standard model with linear production. With either constant savings out of output or the usual discounted preferences with a constant elasticity utility function, this kind of model can be explicitly solved for constant growth rate paths for K and Y .

The model proposed here is essentially one of endogenous technological change. As stated, this technological change operates in a disembodied fashion. An individual firm receives the benefits of growth in the aggregate stock of knowledge E even if it chooses not to invest in additional private knowledge e or capital K . Like the original Solow model, this model could be recast in a form where the effects of increases in the public stock of knowledge can be exploited only if accompanied by additional investment on the part of the individual firm. The details in what follows would change, but not the main conclusions.

Although the issues appear superficially to be related, whether technological change is embodied is quite different from whether the rate of technological change is endogenous. In both the Solow model and this model, the technological change can be put in embodied or disembodied form. In either form, the key feature of the model in this section is that the rate of technological change is increasing in aggregate investment. Thus, an increase in investment has both the usual direct effect on output and an indirect effect on technological change. In contrast, for both the disembodied and embodied versions of the Solow model, the technological change terms $\Gamma(t)$ or $A(t)$ are unaffected by any other variables in the model. Since they are the fundamental sources of growth in that model, it follows almost immediately that there is little scope for policy to affect growth rates¹⁵ or for the conventional aggregate measures of inputs to explain growth rates.

As described in Romer (1986a), this model has a well-defined competitive equilibrium with externalities, despite the presence of increasing returns. As a normative matter, the presence of externalities means that the equilibrium rate of growth for \hat{K} is too small, and that the rental rate on capital differs from its social marginal product. Thus, the elasticity of output with respect to capital differs from the share of capital in total income. Nonetheless, given the approximations underlying equa-

15. For a discussion of the limited role for policy that is inherent in any model of growth based on exogenous technological change, see Rebelo (1987).

tion (5), a growth accounting equation follows by logarithmic differentiation,

$$\hat{Y} = (\alpha + \theta)\hat{K} + (1 - \alpha)\hat{L}, \quad (6)$$

The parameter α can still be estimated as capital's share of total income. According to this model, the conventional growth accounting exercise yields a residual because the weight applied to capital is too small.¹⁶

2.3. INCREASING RETURNS DRIVEN BY SPECIALIZATION

Any departure from the basic growth accounting framework must rely on a departure from perfect competition. The model in the last section does this by introducing external effects and increasing returns. It uses the equilibrium concept of competitive equilibrium with externalities. The model in this section introduces fixed costs and uses the equilibrium concept of monopolistic competition. Details are presented in Romer (1986b, 1987).

The motivation for this model is the old idea that increases in the extent of the market lead to increases in the degree of specialization, and that the increase in specialization has productive value. This productive effect for specialization is captured by using an extension of the "variety in production" model used in Ethier (1982). Suppose that final output depends on labor L and a continuum of possible specialized inputs in production $X(i)$, $X: \mathbb{R}_+ \rightarrow \mathbb{R}$. To describe the basic ideas in the context of a familiar functional form, let

$$Y(X, L) = L^{1-\alpha} \int X(i)^\alpha di. \quad (7)$$

Ignoring the presence of the input L , this is just a continuum version of the Dixit-Stiglitz (1977) functional used to capture a preference for variety in consumption. While it does depend on the variety of inputs used, this form of production exhibits constant returns to scale. Thus, there is no problem assuming that there are many producers in the final goods industry and that they are price takers in the markets for intermediate inputs and for final output. Monopolistic competition will arise only in those industries producing the intermediate goods. By allowing for household production as well as market production, this formulation includes the usual preference for variety as a special case.

Following the classical (not neoclassical) economists, this functional

16. See Lucas (1985) for a model in which positive externalities are associated with labor or human capital instead of physical capital.

form is motivated by the idea that increasing the number of specialized inputs leads to more efficient production. For example, it is possible to connect two pieces of wood with wooden pegs made by hand, but it can be done more efficiently if someone like a blacksmith supplies specialized inputs like hammers and nails. These inputs can themselves be made more efficiently with machinery designed specifically to produce them, and these machines will in turn rely on the availability of other inputs like steel and energy and will themselves be manufactured most efficiently with the help of specialized tools and dies.

As suggested even by Adam Smith, what limits this process of specialization is the extent of the market, that is, fixed costs that must be amortized over a sufficiently high level of output. Thus, assume that the specialized inputs are created from a stock of primary unspecialized capital Z . Assume that $X(i)$ units of any input i can be produced at a cost

$$G(X(i)) = c_0 + c_1 X(i). \quad (8)$$

The feasibility constraint for the economy at time t is

$$\int G(X(i)) di \leq Z(t). \quad (9)$$

Because of the symmetry in the model, the planning optimum for this economy would be one in which the list of specialized inputs takes on the form $X(i) = N/M$ for $i \in [0, M]$ and 0 otherwise. Here N represents the total quantity of inputs, M the range or degree of specialization of these inputs. For this kind of input list, output can be written

$$Y = L^{1-\alpha} M^1 - \alpha N^\alpha. \quad (10)$$

Were it not for the presence of the fixed costs, the solution would be to drive M to infinity, holding N constant; that is to make the degree of specialization arbitrarily large.

Because of the fixed costs involved in the production of any individual good i , only a single producer will produce that particular good. Even in the presence of an infinite number of similar producers offering goods $X(j)$, producer i will still face a downward-sloping demand curve for his good and will therefore have market power. For the simple functional form used here, the inverse demand curve takes the constant elasticity form $p(i) = X(i)^{\alpha-1}$ regardless of the prices charged by suppliers of other intermediate inputs. The firms producing final output goods using the technology described in equation (7) will take as given the price of output goods and a price list $p(i)$ —one price from each intermediate

goods producer—and will choose a list of inputs $X(i)$.¹⁷ Each producer of a specialized input will treat the price of the primary good Z parametrically and will maximize profits taking the demand curve $p(i) = X(i)^{\alpha - 1}$ he faces as given. With free entry into the production of specialized inputs, entry will occur and the price for Z will increase up to the point where equilibrium profits are zero.

For the functional forms chosen here, it is a matter of algebra to show that the static equilibrium for a fixed initial supply of Z is one with $X(i) = \bar{X} = c_0/c_1 [\alpha/(1 - \alpha)]$ on the interval $[0, M]$, where M is determined by $M = Z(1 - \alpha)/(c_0)$. In this case the total quantity of specialized inputs N is given by $N = \bar{X}M = \alpha z/c_1$ and output Y becomes

$$Y = BZL^{1 - \alpha} \quad (11)$$

for a suitable constant B . Increasing returns as a function of labor L and primary capital Z arise because of the interaction of the fixed costs for producing the intermediate inputs and the dependence of final output on the range of specialized inputs.

Now introduce the usual representative individual with standard discounted preferences and a constant elasticity utility function, and suppose that it is possible to accumulate the primary capital good $Z(t)$ like conventional capital,

$$\dot{Z}(t) = Y(t) - C(t). \quad (12)$$

As always, many different equivalent market structures can support the equilibrium. It is simplest to think of consumers accumulating Z directly and renting it to the firms that use it to produce a flow of specialized intermediate inputs.

In another of the many demonstrations of the beauty and simplicity of power functions, it turns out that one can explicitly calculate the dynamic, monopolistically competitive equilibrium for this model. It is not the same as the first best optimum that emerges as the solution to the social planning problem; in fact, it bears a strong formal resemblance to a competitive equilibrium with externalities. The apparent externality (or pecuniary externality if you prefer) arises because an individual producer deciding whether or not to introduce an additional specialized good balances the cost of producing the good against the profit he can extract by selling it at the standard monopoly price. The social benefit

17. Formally, we can allow the price of any intermediate good that is not produced in equilibrium to have a price equal to ∞ .

from the introduction of the good is larger than the profit the monopolist can extract; the difference is the surplus received by the demanders of the good. Thus there is a benefit associated with the introduction of new goods that is not captured by the price system; the model behaves as if there is a positive externality associated with the range of inputs or equivalently with the degree of specialization. This difference between social and private valuations is distinct from, and in this case more important than, the usual wedge between price and marginal cost introduced by monopoly pricing.¹⁸

Given that output Y in equation (11) is linear in the accumulated capital good Z , it is not surprising that both the equilibrium and the planning solution for this economy exhibit constant exponential growth. Given the vague description of why too few goods may be introduced in equilibrium, it should not be too surprising that the equilibrium has a rate of accumulation of Z , and hence a rate of growth of output, that is too small relative to the first best social optimum.

Although the interpretation of this model is quite different from that for the model with knowledge spillovers, the implications of the two models are quite similar. Endogenous, unceasing growth is possible. Growth accounting will fail because of the divergence between the social and private valuations of capital. The share of capital in total income will be α , the share of labor will be $1 - \alpha$, and the equation for the evolution of output is

$$\dot{Y} = \dot{Z} + (1 - \alpha)\dot{L}. \quad (13)$$

Once again, a conventional growth accounting exercise will estimate a residual because it puts too small a weight on the growth rate of primary capital.¹⁹

2.4. LABOR-SAVING INNOVATION

The previous two models suggest reasons why the capital elasticity of output might be higher than capital's share in total income, but in each

18. To demonstrate that these effects are in fact distinct, Romer (1986b) constructs an example where the wedge between price and marginal cost disappears, but where the difference between the social and private valuation of an additional good remains. In this setting, the basic results presented here still go through.
19. In a model with market power but without a dependence on variety, Hall (1986) concludes the opposite, that the weight on capital is too large and the weight on labor too small. His theory is motivated by the familiar observation that production function estimates that rely heavily on cyclical fluctuations—for example, those based on annual or quarterly data—assign an elasticity of 1 to labor. The interpretation of these empirical results is discussed below.

case the labor elasticity is still equal to labor's share in total income. The theory outlined in this section offers a separate explanation for why the elasticity with respect to labor might also differ from its share in income. As described in the introduction, the basic idea is that some forms of knowledge may be substitutes for labor rather than complements. That is, these forms of knowledge are labor saving in the strong sense that at a constant wage rate, an increase in knowledge leads to a reduction in the quantity of labor demanded.²⁰

The empirical problem that this kind of theory faces is obvious. To have any hope of explaining the trend in wages, this kind of effect must be augmented by some other explanation for rising wages. For illustrative purposes, I will graft a version of endogenous development of labor-saving innovation onto the model of specialized goods from the last section, but this choice is somewhat arbitrary.

To formalize what a model of labor-saving innovation might look like, let e stand once again for the private stock of knowledge held by a representative final goods producing firm, and let E be the aggregate stock of public knowledge that results from spillovers from the private stock. Suppose for simplicity that labor now enters final output as part of a composite good H that itself depends on a composite knowledge variable, $\psi(e, E)$. Thus equation (7) becomes

$$Y(L, X, e, E) = H(\psi(e, E), L) \int X(i)^\alpha di. \quad (14)$$

Solving the static monopolistic competition model as before for a given initial stock of primary capital Z , this becomes

$$Y = AH(\psi, L)Z. \quad (15)$$

One of the easiest ways to represent labor-saving innovation is to allow knowledge ψ to act as a perfect substitute for L , for example to represent robots made of silicon chips that are costly to design, yet essentially costless to produce. Once designed, they are a perfect substitute for workers. Then the function H could take the form

$$H(\psi, L) = (\psi + L)^{1 - \alpha}. \quad (16)$$

For ψ to be labor-saving in the strong sense identified above, all that is required is that the cross partial derivative of H is negative. To ensure

20. No elaboration of the model outlined in this section is available, but a closely related model that relies on technological change that is resource-saving in the same strong sense is presented in Romer and Sasaki (1986).

that diminishing returns do not force research to stop, ψ would need to exhibit increasing returns. This poses no problem for equilibrium provided that $\psi(e, E)$ is concave in e when E is held fixed.

The problem for this economy at any point in time is to allocate current output between consumption, accumulation of new capital Z , and investment in the production of new knowledge e . In principle, once one specifies intertemporal preferences and accumulation functions for e and Z and the relation between private knowledge e and public knowledge E , it is straightforward to write down the dynamic equations that will characterize the equilibrium, even though it involves both market power and externalities. However, there is no hope that a model using a function like H in equation (15) can be solved explicitly for paths for $e(t)$ and $Z(t)$.

Some of the properties of this kind of model can still be inferred. The relevant observation is that holding $Z(0)$ and $e(0)$ constant, an increase in $L(0)$ will reduce wages and increase the return to investment in primary capital Z —the usual effects—but will also decrease returns to investment in private knowledge e . Thus, exogenous growth in L could reduce the rate of investment in e . If the spillovers from e to E are important enough, an increase in the growth rate of L might cause the composite $H(\psi(e, E), L)$ to grow more slowly, which would in turn induce slower growth in Z .

Innovation is a good from the point of view of society as a whole. Because of the externalities associated with spillovers of knowledge, investment in labor-saving innovation is suboptimal in equilibrium, so any effect that influences this investment can have large welfare effects, not just little triangles. Because it is a second-best equilibrium, these welfare effects may be perverse. In the extreme case where a reduction in the rate of growth of labor stimulates an increase in the growth of e and E large enough to cause a net increase in the growth of $H(\psi(e, E), L)$, a policy that forces up wages and the costs of employment so much that a nontrivial fraction of the existing work force becomes unemployed might actually have a positive effect on the rate of productivity or even output growth. Of course, it is never fully optimal to have wasted resources. A better intervention would be to support innovation directly.

To derive a growth accounting equation for this economy, take a logarithmic approximation to equation (15), and write

$$\hat{Y} = \hat{Z} + \beta_1 \hat{e} + \beta_2 \hat{E} + \beta_3 \hat{L}, \quad (17)$$

where β_i represents the derivative of $\ln H(\psi(e, E), L)$ with respect to the logarithm of e , E , or L . Suppose that the growth rate of E is positively

related to the growth rate of e (for example, that it moves one-for-one with e) and that the growth rate of e responds negatively to changes in the growth rate of L because of the interaction through wages. Then the coefficient on the growth rate of labor in a regression that relies only on the observables \hat{Z} and \hat{L} will equal the sum of β_3 and a negative coefficient times $\beta_1 + \beta_2$. Even in a growth accounting exercise that correctly measured \hat{e} (for example by making use of data on patents or research expenditure), the estimated coefficient on \hat{L} would still contain a negative effect due to spillovers captured in E .

3. Evidence

There is a great deal of evidence that bears on the issues raised in the last section. It is organized in this section as follows. Section 3.1 considers informal direct evidence about the nature of production. This kind of evidence is not usually presented explicitly, but seems nonetheless to play an important role in our willingness to consider different theoretical models. Consider, for example, a model in which conventional physical investment by one firm is assumed to have direct external effects on the technologies of other firms. Most economists would not be willing to let this kind of model even get near a body of data because it runs counter to all of our experience and understanding of the world. This first section is designed to convince a skeptical reader that the alternative models considered in section 2 are worthy of an encounter with the data.

Ultimately, all the questions about the behavior of productivity reduce to a question about the joint movements of capital and labor and about elasticity of output with respect to these inputs. Section 3.2 considers evidence from the United States on these issues, starting from a long historical perspective at the aggregate level and working toward the present and a more disaggregated level. One of the main conclusions to emerge is that long time intervals may be necessary to disentangle the familiar high-frequency fluctuations associated with business cycles from the low-frequency movements considered here. Thus, the entire historical record for the United States effectively offers only a few degrees of freedom for the relevant estimates.

All of the evidence considered for the United States could in principle be repeated for other countries. To keep the task attempted here manageable, only a partial extension along these lines is attempted. In this paper, the only evidence from other countries that considers productivity at a disaggregated level relies on data collected by Robert Gordon (described in Gordon 1985, 1986). He has collected wage and labor productivity data for three regions, the United States, Japan, and an aggre-

gate of European countries, all for the interval from 1961 to 1984. These data have the advantage of distinguishing between manufacturing and nonmanufacturing, but do not contain comparable capital stock estimates. They are considered in section 3.3.

Returning to the aggregate level, there are two ways to expand the available data: using longer time intervals or more countries. Because of problems with the availability of data, these directions are mutually exclusive and must be undertaken separately. Section 3.4 contains two separate preliminary attempts in these directions, considering long-run evidence from several countries using data from Maddison (1982), and cross-sectional evidence from the largest possible collection of countries using data after 1960 from Summers and Heston (1984).

3.1. INFORMAL EVIDENCE ON THE PLAUSIBILITY OF THE ALTERNATIVE MODELS

The intuitive issues surrounding the Solow model are quite familiar and need not be reviewed here. The idea that there are spillovers of knowledge is apparently not controversial, and estimates from case studies suggest that they can be fairly large.²¹

The idea that the rate of increase of productive knowledge is closely tied to the rate of investment in capital is more controversial. If by knowledge one means progress in basic science, then surely this is wrong; but basic science is probably not the appropriate index of productive knowledge. For example, Nathan Rosenberg (1976, chap. 15) points out that neither England nor the United States was a major center of basic science at the time when each rose to a position of world prominence in production. Similarly, the remarkable industrial performance of Japan in the postwar era has not been matched by achievement in basic science.

Using long time series on patents and investment from four major industries—petroleum refining, papermaking, railroads, and farming—Jacob Schmookler (1966) presents convincing evidence that invention is in fact very closely associated with capital investment and that inventive activity does indeed respond to the same kinds of economic incentives that drive investment. To cite just two small parts of the large body of evidence that he presents, Schmookler finds that there is a strong time-series correlation between rates of patenting in areas that face different technological opportunities but similar economic opportunities—for example, track and nontrack patents for railroads, or sole-making and leather-sewing patents for shoe manufacturing. More directly, there is a strong correlation between the level of patenting and the level of invest-

21. See for example the estimates for the chemical industry in Bernstein and Nadiri (1986).

ment both over time and across industries. Patenting tends to follow investment, and the correlation with investment is stronger than that with either the level of output in the industry or the general level of economic activity.

Schmookler also points out that the much-noted pattern whereby inventive activity in an industry follows a logistic time trend may be due to declines in the economic value of patents rather than a decline in technological opportunities. Thus he concludes that patenting fell off in railroads because of the arrival of highways and automobiles, not because of technological depletion. He also cites the case of patents on horseshoes, a simple piece of technology that dates from the second century. Patentable improvements in horseshoes continued in significant numbers up until the 1920s, and peaked around 1900. If strong demand can induce improvements in a 1700-year-old technology as simple as this, I find it incredible that we have now exhausted the opportunities for technological improvement in areas like steam-electric generation or chemical processing. Although the authors of careful case studies of these areas would probably disagree (see Gordon 1983, and Baily and Chakrabarti 1985, respectively), I suspect that a sharp reduction in demand growth and a consequent reduction in the demand for capital investment may be largely to blame in each case. (Note, incidentally, that while the laws of thermodynamics place an absolute limit on the output of electricity per unit of coal, they place no obvious limits on output per manhour or per unit of capital.)

The informal evidence supporting a productive role for specialization is pervasive. For example, the availability of specialized goods offers a specific foundation for the vague agglomeration effects used to explain the formation of cities. Producers and consumers may locate in cities not so much to take advantage of external effects in the usual sense, but rather to take advantage of goods and services that are not available elsewhere. The disproportionate growth of cities like Mexico City may seem mysterious until one considers the prospect of locating a small firm—for example, an economics department—somewhere in Mexico. In how many other locations could one hope to purchase same-day service on the photocopy machine, let alone the PC's?

Wallis and North (1986) also provide evidence that is relevant on this point. They provide estimates of the fraction of GNP that is devoted to the execution of transactions. This involves estimating the resources used in the "transaction" industries, such as wholesale and retail trade, and banking, insurance, and real estate. It also involves estimates of the costs of providing transaction services in the nontransaction industries, for example the costs of compensation to individuals who are engaged in

purchasing inputs or supervising others. Their overall estimate is that the transactions component has increased from 25 percent of GNP in 1870 to 45 percent in 1970. These kinds of figures are hard to explain without recourse to some model with a growing number of specialized goods.

The idea that the level or rate of growth of wages might influence the rate of innovation has been used to explain broad features of the historical behavior of productivity in the United States. First Rothbart (1946) and Habakkuk (1962) argued that the availability of land in the United States raised wages relative to those in Britain, and that this caused the more rapid productivity growth observed in the United States. The basic facts on which they and subsequent authors seem to agree are that both wages and interest rates were higher in the United States, and that in certain notable cases the level of labor productivity in the United States exceeded that in Britain. Most of the subsequent discussion of this issue has focused on the theoretical assertion that an abundance of land is capable of generating these results in a conventional model of factor substitution under constant returns to scale. Skepticism about the possibility of this result in a simple 2-sector, 3-factor model was expressed by Temin (1966); the most recent contribution to this debate is that the results might arise in a model with 3 sectors and 4 factors (James and Skinner 1985). Relatively little theoretical attention has apparently been devoted to the suggestion made above, and described in verbal terms by Habakkuk (1962, p. 49), that higher wages may have operated through the incentive to invest in innovation and invention, and that these lead to spillovers of knowledge.

Kendrick (1961) identifies an increase in the trend rate of productivity growth in the years immediately following World War I. This has also been interpreted as a response to labor shortages and increased wages, in this case brought on by restrictions on immigration.²² Generalizing from labor shortages to energy shortages, Maddison (1982, p. 34) identifies two instances of what he labels the challenge-response mechanism. Both of the early leaders in productivity, the Netherlands up to 1785 and England from 1785 to 1890, had meager supplies of wood, the primary energy input of the time. The Netherlands responded by using peat and the English developed coal, in each case allegedly receiving a net benefit because of the shortage of wood.

3.2. PRODUCTIVITY AND GROWTH IN THE UNITED STATES

Basic long-run data on productivity trends in the United States are illustrated in figure 1. Average annual growth rates for output per hour

22. See for example Oshima (1984).

worked and for the labor force are given for successive twenty-year intervals.²³ If the last observation were split into the separate decades of the 1960s and 1970s, the magnitude of the recent slowdown would be more pronounced. The striking feature about this graph is the negative correlation between the rate of growth of the labor force and the rate of growth of output per hour worked. Since the growth of output per hour worked equals the growth of output minus the growth in the labor force (and in years after 1869, plus growth in unemployment and minus

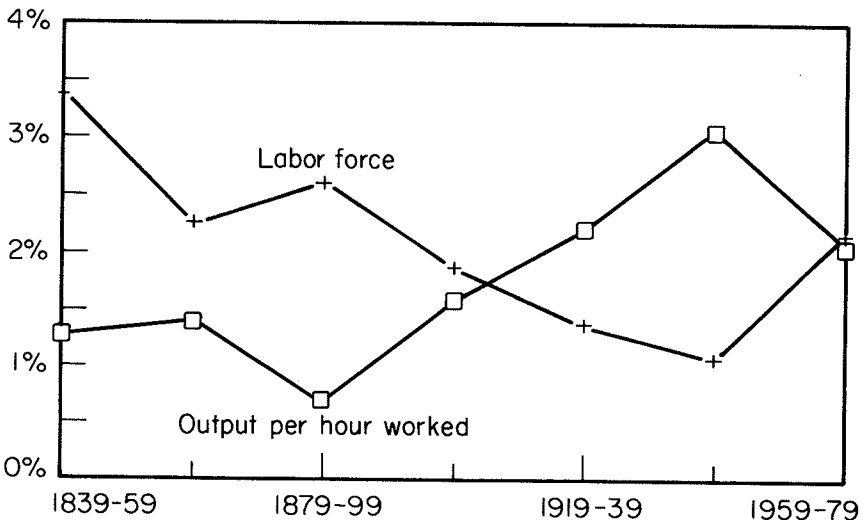
23. Data for figure 1 were gathered from the following sources.

Output growth: 1839 to 1859, Gallman (1966); 1859 to 1869, Kuznets (1971); 1869 to 1889, Friedman and Schwartz (1982); 1889 to 1949, total private GNP, table A.III, Kendrick (1961); 1949 to 1979, Bureau of Labor Statistics, business sector output, *Monthly Labor Review*, December 1983, 1986. Gallman collected data on an annual basis, but did not report it in a form that can be used to calculate growth rates for the decades after 1859. Friedman and Schwartz had access to his annual worksheets and report those numbers starting in 1869. Presumably Kuznets, who was Gallman's thesis supervisor, did as well. For the years in which they overlap, Kuznets' numbers are close to those reported by Gallman. After 1889, a measure of output that excludes the government is used because government output is measured in terms of labor input, so productivity figures for this sector are meaningless.

Labor force: 1840 to 1930, Lebergott (1966); 1929 to 1979, *Economic Report of the President*, 1984.

Hours worked: for 1840 to 1870, growth in hours worked is assumed to be the same as growth in the labor force because no independent measures of unemployment and average hours are available. From 1869 to 1949, private sector hours are taken from table A.X, A.XI, Kendrick (1961). From 1949 to 1979, hours are from the BLS source cited above.

Figure 1 PRODUCTIVITY AND LABOR FORCE MOVEMENTS



growth in average hours worked), errors in measurement of growth in the labor force could be partially responsible for this negative correlation. However, labor force numbers based on census data should be a relatively accurate historical series.²⁴ Moreover, an error on the order of 20 percent in the level of the labor force in a particular year would be needed to induce a 1 percent error in a twenty-year growth rate.

Labor force movements are used here as an indication of exogenous changes in labor supply. Although immigration in the early years may in fact have varied somewhat with the growth of output, a labor force measure is presumably less sensitive than either unemployment or hours worked to business cycle fluctuations. Since output and hours worked move together almost one-for-one over business cycle fluctuations, the negative correlation evident in this figure would be weaker if hours worked had been used instead of the labor force.

For an economy on a steady-state growth path, the Solow model implies that changes in the growth of the labor force should induce one-for-one changes in the growth of output. Thus, if twenty-year data were sufficient to represent steady-state growth (free of business cycle fluctuations), output per hour worked would be independent of growth in the labor force. In fact, what the data suggest is that labor productivity responds quite negatively to increases in the labor force. If the assump-

24. For example, Weiss (1986) presents revisions of the Lebergott estimates of the level of the labor force for the years 1840, 1850, and 1860 that differ from the original estimates by less than 1 percent.

Table 1 PRODUCTION FUNCTION ESTIMATES, DECADE DATA, 1890s TO 1970s

	Exogenous technological change	<i>Elasticity of output with respect to</i>		
		Capital	Hours worked	Labor force
1.		1.01 (.29)	.32 (.53)	
2.		.95 (.39)		.38 (.63)
3.	2.0 (.6)	.18 (.31)	.76 (.37)	
4.	2.1 (.9)	.87 (.31)		-.64 (.66)

Note: Standard errors in parentheses. The basic equation is a regression of output growth rates on capital and one of two measures of growth rates for labor supply. Units are average annual (continuously compounded) percentage growth rates times 100, so the constant terms have units of percent per year. For a description of the data, see the data footnote to figure 1.

tion of steady-state growth is dropped, some negative relation would be expected. For example, if $Y = \Gamma(t)K^\alpha L^\beta$ with $\alpha = .33$ and $\beta = .67$, and if changes in the rate of growth of K and Γ are absent or at least uncorrelated with changes in L , one would expect a regression of $\hat{Y} - \hat{L}$ on \hat{L} to yield a coefficient of $\beta - 1 = -0.33$. In fact, if we use the non-parametric Theil estimate for the slope of this relation in these data, the estimate is $-.93$ with a 93 percent confidence interval of $(-1.51, -.46)$.²⁵

Table 1 extends this observation with the use of capital stock data from Kendrick (1961, table A.XV) starting in 1889. Even with data measured over a single decade rather than two decades, this permits only a meager 9 data points. The table reports regression estimates of growth of output on a constant, capital stock growth, and growth in two measures of labor, the labor force and hours worked.

The conclusion that emerges from the table is that allowing for the simultaneous effects of capital and labor does not change the small estimated coefficient on labor unless one is willing to allow for exogenous technological change (that is, a constant term in the log difference equations), and use hours worked rather than the labor force as a measure of labor. The labor force measure is a worse measure of actual labor input, but one that is symmetric with the measure used for capital. Even using decade data, the cyclic variation of hours worked relative to the size of the labor force is evident, especially during the 1930s.

The results in regression equation 3 in table 1 are similar to results typically reported for production estimates using annual data—a coefficient on capital that is insignificant and a coefficient on labor that is large and significant. Table 2 reports a variety of production function estimates for annual data from 1950 to 1984 for the private business sector. The output, capital stock, and labor indices are those used by the BLS to construct the official estimates of multifactor productivity. The capital stock measure is a sensible Tornqvist aggregate and the output, capital, and hours measures are constructed to have the same coverage. (See Mark and Waldorf 1983, and Mark 1986, for details.)

Regression equation 1 in table 2 yields results similar to those from table 1. If no exogenous trend is allowed, the coefficient on capital is close to 1 because the capital-output ratio has remained roughly constant. (Since

25. The Theil estimate for the parameter b in an equation of the form $y = a + bx + \varepsilon$ is the median of the ratios of the form $(y_i - y_j)/(x_i - x_j)$. The estimator is distribution-free in the sense that the confidence interval for b is exact in finite samples regardless of the distribution of the error terms. All that is required is that the errors be independent and symmetric. It is possible to calculate exact confidence intervals only for specific values, hence the use of a 93 percent interval instead of the usual 95 percent interval. For details, see Hollander and Wolfe (1973).

all variables have been transformed by taking logarithms, these coefficients can be interpreted as elasticities.) As is clear from the subsequent estimates, this equation imposes restrictions that are not consistent with the year-to-year fluctuations in capital, output, and hours worked. Equation 5 reports a more familiar production function type estimate using annual data, a regression in differenced form of output growth rates on growth rates of capital and hours worked. This reproduces the usual result that the coefficient on hours worked is close to 1 and that on capital insignificant (and negative), but it too imposes restrictions that the data

Table 2 ANNUAL DATA, 1950-1984, PRIVATE BUSINESS SECTOR

A. Dependent variable: <i>Y</i>								
	<i>Trend</i>	<i>Y</i> (-1)	<i>K</i>	<i>K</i> (-1)	<i>H</i>	<i>H</i> (-1)	<i>DW</i>	<i>SSR</i>
1.			.95 (.08)		-.15 (.27)		.55	.080
2.	.044 (.010)		-.48 (.34)		.42 (.25)		.34	.050
3.	.028 (.007)	.71 (.11)	-.70 (.22)		.53 (.17)		1.39	.021
4.	.009 (.003)	.60 (.06)	1.47 (.36)	-1.28 (.31)	.83 (.08)	-1.18 (.08)	2.30	.002
5.	.042 (.012)	1.0*	-.60 (.34)	.60*	1.10 (.12)	-1.10*	1.07	.010

*Restrictions imposed by using differenced data for this regression.

B. Dependent variable: <i>K</i>						
	<i>K</i> (-1)	<i>Y</i>	<i>H</i> (-1)	<i>AR</i> 1	<i>DW</i>	<i>SSR</i>
6.	.83 (.02)	.14 (.02)	.20 (.03)	.63** (.14)	1.73	.0003

**Estimated using iterative Cochrane-Orcutt.

C. Dependent variable: <i>H</i>								
	<i>Trend</i>	<i>H</i> (-1)	<i>Y</i>	<i>Y</i> (-1)	<i>K</i>	<i>K</i> (-1)	<i>DW</i>	<i>SSR</i>
7.	-.007 (.003)	1.21 (.11)	.94 (.09)	-.65 (.06)	-.83 (.46)	.74 (.40)	2.4	.0028

Note: Standard errors are given in parentheses. *DW* is the Durbin-Watson statistic, *SSR* is the sum of squared residuals. Trend is a linear time trend. Other variables are the logarithms of the levels of output (*Y*), capital (*K*), and hours worked (*H*). The equations in panels B and C are the result of a mechanical procedure of excluding variables with *t* statistics less than 2, and, in the case of panel C, allowing for an *AR*(1) disturbance term to remove the serial correlation in the residuals. As noted in the text, the structural interpretation for all these equations is unclear.

reject. The F statistic for a test of these restrictions is 36, with a 1% significance level of 4.5.

It should not be surprising that production function regressions using annual data yield estimates that are ambiguous. They cannot be justified in the context of a model of business cycles that allows for productivity shocks, serially correlated disturbances to output, or contemporaneous feedback from output to investment and employment within a year. One might just as well use capital or hours worked as the dependent variable (as in panels B and C) and interpret the regression as an investment equation or an employment equation.

The difficulties in interpreting these kinds of estimates and giving them a structural interpretation are well known. There is substantial disagreement about the fundamental shocks that drive annual macroeconomic data, so assertions about exogeneity are tenuous at best. Both the stock of capital and hours of labor used may be subject to costs of adjustment, so measured inputs may not accurately reflect true inputs. Moreover, the signal-to-noise ratio may simply be too low and the number of observations too few to permit any valid inferences about low-frequency questions. Indicative of this last difficulty is the fact that despite the entrance of the baby-boom cohort into the labor force and the changes in labor force participation by women during the 1970s, the size of the labor force does not help to explain the behavior of hours worked. The labor force variable is not significant in equation 7, and multicollinearity does not seem to be the problem. Excluding the trend, the capital stock variables, or both does not help.

Given these difficulties, there is very little that can be learned about production functions using data like a post-war annual time series for a single country. In principle, the problems of simultaneity can be overcome by using instrumental variables, but there is little hope that any valid instruments exist. For long time series, filtering out business cycle frequencies by averaging over long intervals may mitigate, but not entirely remove, these problems.

These issues aside, the key difficulty with the official BLS data used here is that they do not yet make an adjustment for changes in the quality of labor; however, preliminary estimates of the effects of changes in the sex ratio and levels of education and experience are available (Waldorf, Kunze, Rosenblum, and Tannen 1986). Adjustments are made on the basis of wage equations estimated using both the usual Census data and detailed Social Security data on experience. Annual data for quality adjustment are not made available, but the authors do present a table showing that this adjustment increases effective labor force growth on average by 0.26 percent per year from 1948 to 1973, 0.03 percent from

1973 to 1979, and 0.48 percent from 1979 to 1984. Data reflecting these adjustments are presented in table 3. Qlabor denotes labor hours corrected for changes in quality.

Table 3 shows that if one splits the sample into equal thirds, the quality adjustment makes very little difference. It also shows why one may be forced into the view that something other than capital-related externalities and increasing returns are at work here. The rate of growth of the capital stock does not vary much over this period. There is a fall in output per unit of capital in the last period that arises primarily from the fall in the rate of growth of output. This may partly reflect the cyclical downturns in 1975 and 1979–1982 that are contained in this period.²⁶ More striking is the fall in labor productivity, measured either with or without a correction for quality. Consistent with the long-run evidence presented in figure 1, this fall in output per hour worked is the result of an increase in hours worked that is not accompanied by an increase in output.

Figure 2 charts the behavior of labor productivity in manufacturing using the BLS data. It plots the logarithm of labor productivity and the predicted values based on a regression with a constant and a linear trend. It documents the observation made in the introduction that there is no evidence whatsoever of a secular slowdown in labor productivity in this sector. Consistent with the emphasis placed here on the rate of growth of labor, this sector did not experience an increase in the growth

26. I have intentionally divided the interval here into three equal parts rather than follow the usual practice of measuring rates of change from business cycle peak-to-peak. This practice would be fine if business cycles were like sine waves. In fact, the choice of peaks is subjective—is 1982 a peak or not? The peak-to-peak method tends to exaggerate the slowdown in the last interval measured, which is almost always not itself a peak. It may offer the misleading impression that the statistics so generated are not subject to stochastic variation because of business cycle fluctuations. Finally, Darby (1984) has argued that price controls make the magnitude of the peak in 1973 suspect.

Table 3 AVERAGE ANNUAL PERCENTAGE RATES OF CHANGE

	1948–84	1948–60	1960–72	1972–84
Private business sector				
Output	3.2	3.1	3.8	2.6
Inputs: Qlabor	1.2	0.4	1.3	1.8
Labor	0.9	0.1	1.0	1.6
Capital	3.5	3.1	3.7	3.5
Output per unit Qlabor	2.0	2.8	2.5	0.7
Output per unit labor	2.3	3.0	2.8	1.0
Output per unit capital	-0.3	0.0	0.1	-1.0

Note: Qlabor denotes labor adjusted for changes in quality.

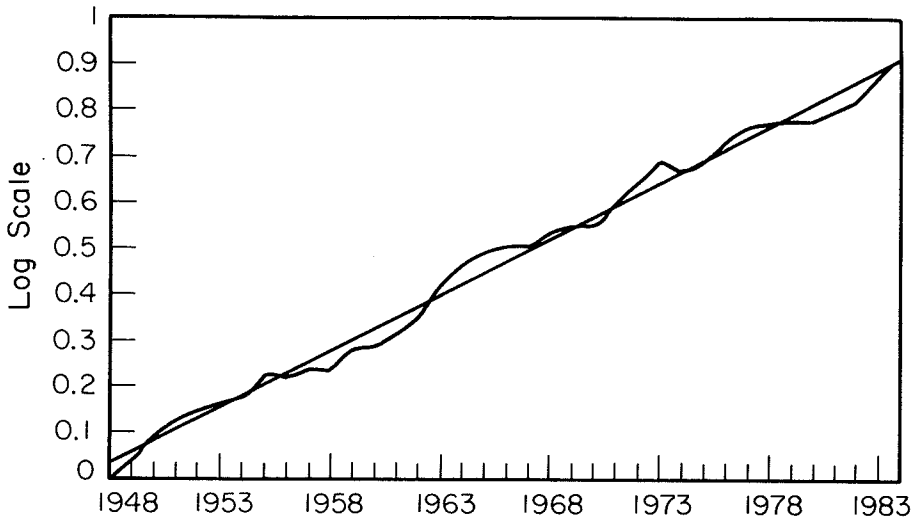
of hours worked in the most recent period. The increase was concentrated entirely in nonmanufacturing.

Previous authors have noted that the three sectors in nonmanufacturing that experienced the largest slowdown in labor productivity are construction, mining, and utilities.²⁷ However, since these sectors account for only 13 percent of private business sector GNP by 1984, much of the explanation for the slowdown must lie elsewhere. Figure 3 plots the logarithm of the level of output, full-time-equivalent workers, and productivity for an aggregate service sector consisting of wholesale trade, retail trade, FIRE (financial institutions, insurance, and real estate), and the NIA service sector,²⁸ with the initial value set in each case equal to zero. Thus, slopes in the graph can be interpreted as rates of change. These sectors accounted for 41 percent of private business sector GNP in 1948 and 52 percent in 1984. The previously noted correlation between the labor productivity slowdown and the increased growth in the labor force now stands out unmistakably. This also suggests a substantial revision in the estimated start of the slowdown, pushing it back to 1964. Taking 1964 as a break point, the 1.7 percent slowdown in the annual labor productivity growth rate can be attributed to a 1.5 percent increase in the annual rate of growth of employment (from 2 percent to 3.5 per-

27. See Allen (1985), Nordhaus (1982), and Gordon (1983) for discussions of these sectors.

28. The data are taken from National Income Account tables 6.2 and 6.7B.

Figure 2 LABOR PRODUCTIVITY IN MANUFACTURING



cent per year) and a fall of 0.2 percent in the rate of growth of output. At most, only a part of the failure of output growth to keep up with employment growth can be attributed to cyclical effects. In a comparison of the intervals 1948–1964 and 1964–1973, a slowdown of 1.2 percent can be attributed to an increase in the growth rate of employment of 1.5 percent and an increase in the rate of growth of output of 0.3 percent.²⁹

3.3. RECENT EVIDENCE FROM EUROPE AND JAPAN

Figure 4 presents data on wages and on output per hour for an aggregate of European countries. Like figure 3, it plots the logarithm of the levels of the variables, each with an initial value set to zero. There is evidence of a fall in the rate of labor productivity growth in both manufacturing and nonmanufacturing, which presumably is at least partially the result of macroeconomic performance in the years after the oil shock; nonetheless, the rates of productivity growth remain high by U.S. standards. For example, from 1973 to 1984, productivity growth in non-manufacturing averaged 2.2 percent per year; in manufacturing it was

29. As noted above, Darby (1984) argues that the price controls lead to an underestimate of inflation in 1973 and an overestimate of productivity and output growth. In the comparison here, using 1973 gives an upper bound to the possible effects of business cycles, allocating as much of the slowdown as possible to the cyclical effects. If Darby is right, the allowance for cyclical effects given in the text is an overestimate.

Figure 3 WHOLESALE, RETAIL, SERVICES, FIRE

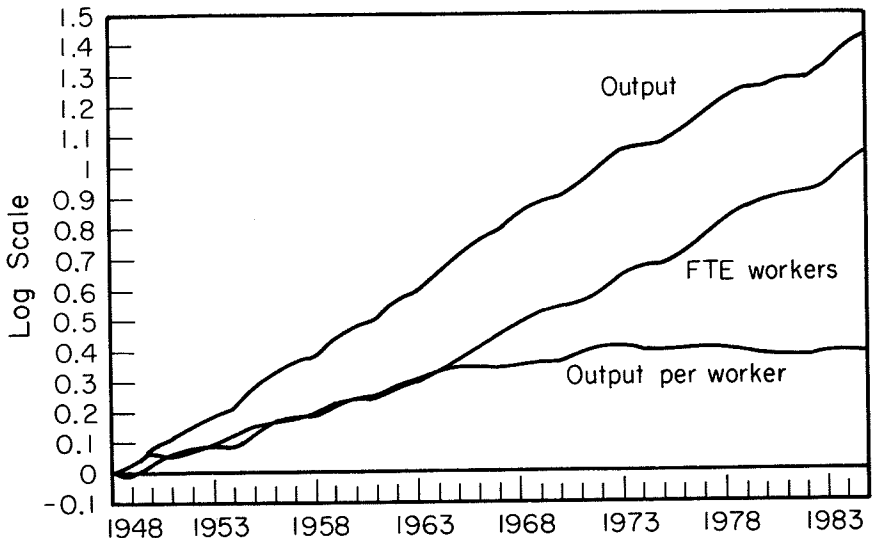


Figure 4 EUROPE: WAGES, OUTPUT PER HOUR

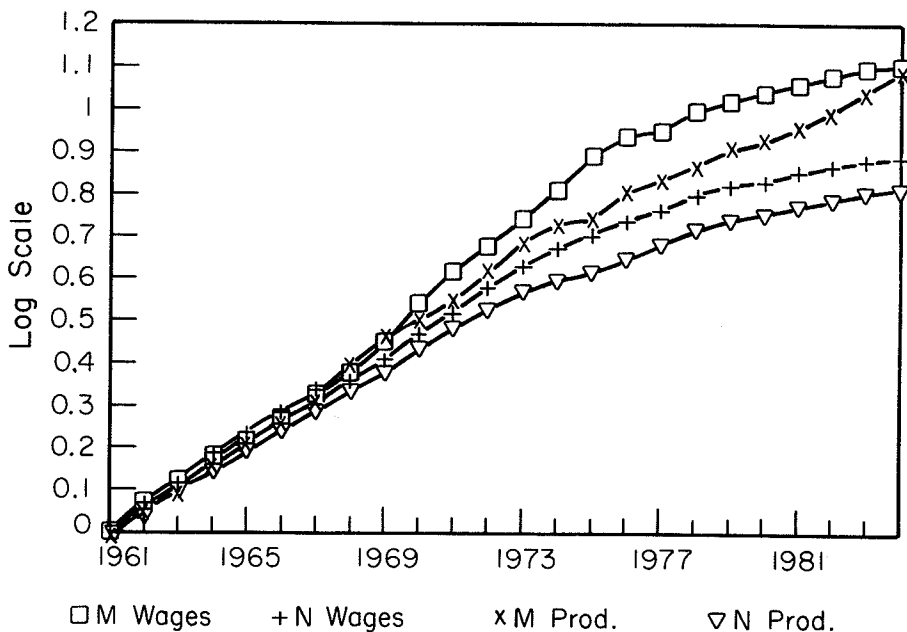


Figure 5 JAPAN: WAGES, OUTPUT PER HOUR

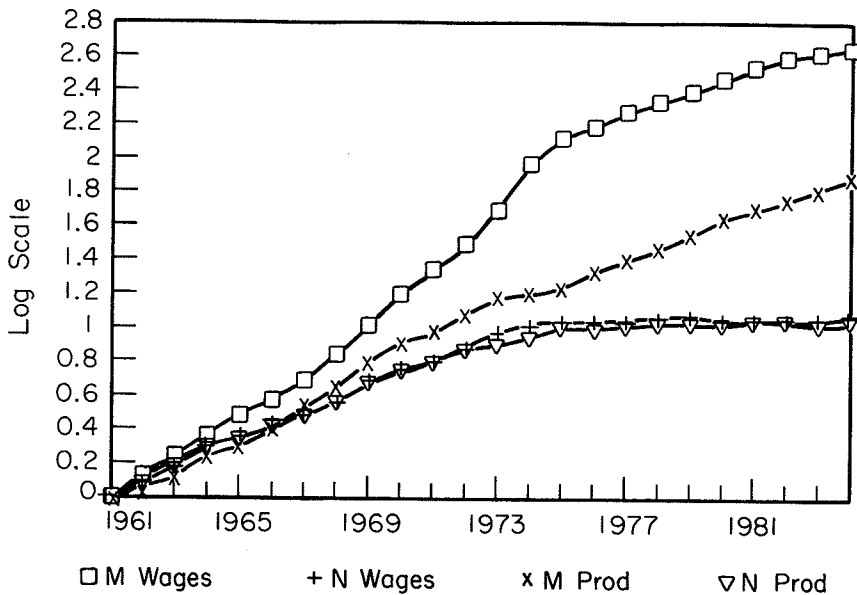


Figure 6 MANUFACTURING TOTAL HOURS

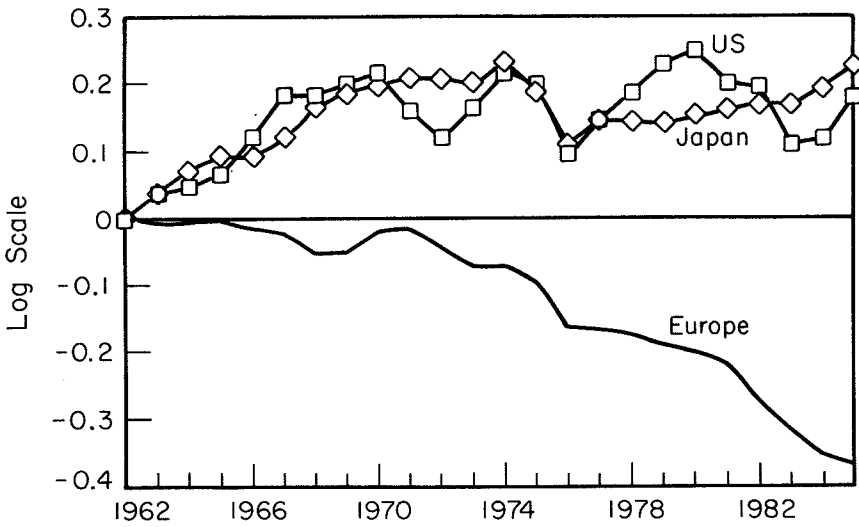
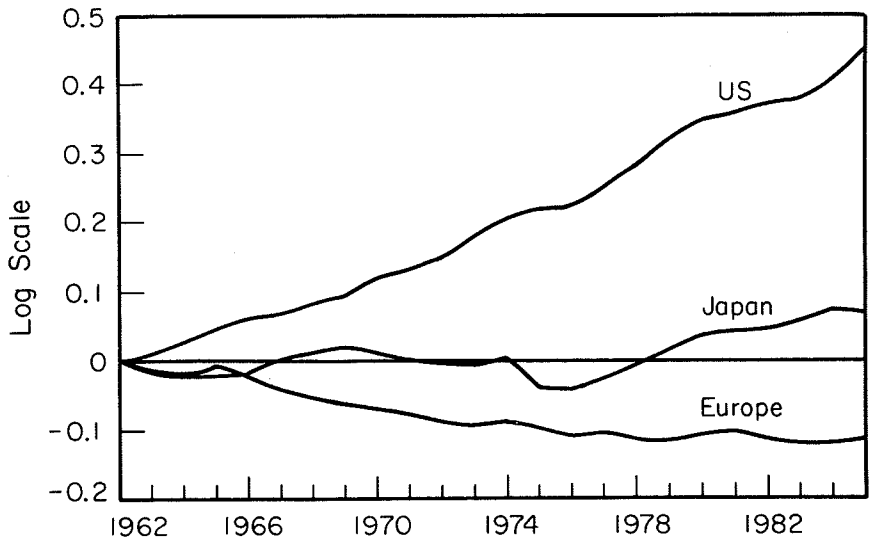


Figure 7 NONMANUFACTURING TOTAL HOURS



3.7 percent per year. Relative to the United States, wage growth in both sectors has continued to be robust.

Figure 5 gives similar results for Japan. Note that the vertical scale here (measured in logarithms) is 2.5 times that for Europe. The evidence of a productivity slowdown in nonmanufacturing after 1973 for Japan is more pronounced than for Europe, but growth in this sector still averaged 1.1 percent per year in the post-1973 period. Productivity growth in manufacturing over this period averaged 6.4 percent.

Figures 6 and 7 compare the growth in total hours in manufacturing and nonmanufacturing respectively for the three regions. In general the negative correlation between productivity and wages on the one hand and growth in total hours on the other is confirmed. Hours have grown relatively quickly in nonmanufacturing in the United States for the entire period and in Japan since 1975. Hours have fallen in Europe in both sectors. The negative association between hours growth and productivity growth fails only for Japanese manufacturing. For example, both were higher before 1967 than after 1974.

The behavior of manufacturing productivity in the early period may reflect the extremely high rates of capital accumulation that took place in Japan in the 1960s. Without detailed sector-specific capital stock estimates this cannot be verified conclusively, but the following evidence is suggestive. Maddison (1982, table D11) reports that the stock of net non-residential, fixed tangible capital increased at the rate of 11 percent per year from 1960 to 1973. From 1973 to 1978 it grew at the rate of 5 percent. If these rates are representative of rates of accumulation in manufacturing, this might be sufficient to explain the growth in wages in the early period, especially if the output elasticity of capital is truly on the order of 1 instead of 0.25.

3.4. LONGER TIME INTERVALS AND MORE COUNTRIES

At the aggregate level the rough growth accounting calculation suggested for Japan in the last paragraph can be carried out formally and extended to other countries. Data from Maddison on cross country rates of growth and accumulation of capital are presented in figure 8. Letters refer to individual countries as follows: C, Canada; E, United States; F, France; G, Germany; I, Italy; J, Japan; K, United Kingdom. The numbers refer to the following time periods: 1, 1770–1870; 2, 1870–1913; 3, 1913–1950; 4, 1950–1979. The sample is limited by the availability of capital stock data. The results from section 3.2 suggesting that capital and output move together one-for-one over long time intervals is strongly confirmed for this set of countries.

Regression results show that allowing for variation in hours does not

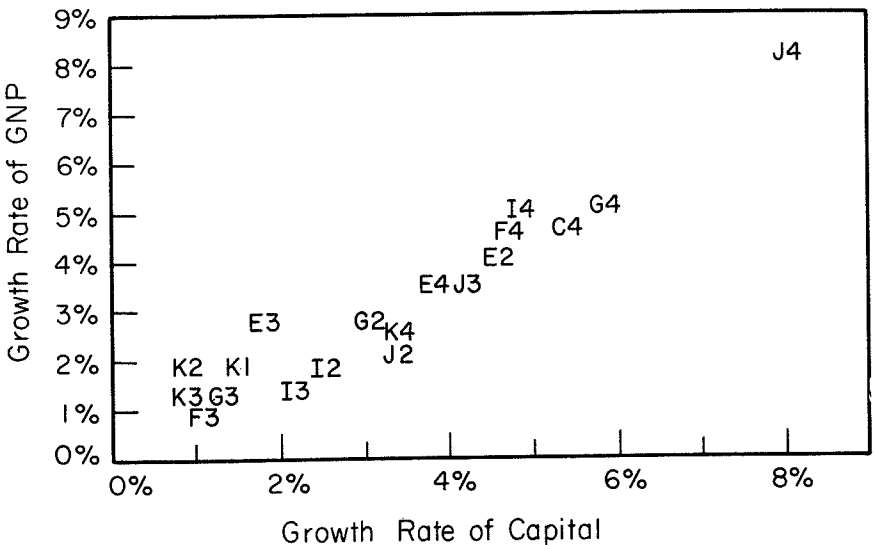
change this impression. Equation (18) reports the result of regressing the observed growth rate for output, \hat{Y} , on a constant, the growth rate of capital, \hat{K} , and the growth rate for hours, \hat{H} . Standard errors are given in parentheses; σ reports the standard error of the regression. The basic data are in average annual growth rates, times 100, so the standard error σ has the interpretation of 0.6 percent per year.

$$\hat{Y} = 0.2 + .87 \hat{K} + .04 \hat{H}, \sigma = 0.6 \quad (18)$$

(.003) (.08) (.18)

It is useful to recall the theoretical results from section 2 at this point and consider what assumptions are needed to explain these results using the neoclassical model. These results are not consistent with the usual explanation whereby countries with low levels of capital per worker grow faster as they accumulate capital and approach the steady-state growth path. As shown by equation 3 from section 2.1, this explanation implies that the regression coefficient of output growth on capital growth, which is essentially the same as the slope in figure 8, should be α , on the order of $\frac{1}{4}$ to $\frac{1}{3}$. In fact it is close to 1. The calculations given in section 2.1 also show that the kind of mismeasurement of quality change inherent in a model of embodied technological change is unlikely to be large enough to explain these results.

Figure 8 CAPITAL AND GNP GROWTH



The only way to explain these results is to suggest that those countries that grew fast were those with high rates of exogenous technological change. One would expect a higher rate of technical change γ , to be accompanied by a higher rate of capital accumulation under the steady-state result described in section 2.1, $\dot{K} = \dot{Y} - \dot{L} = \gamma/(1 - \alpha)$. However, the theoretical results of Sato (1963) suggest that the adjustment of \dot{K} to changes in γ might be slow, and the data from the United States cast doubt on the idea that capital accumulation rates responded much at all to exogenous movements in \dot{L} over twenty-year intervals. An explanation based on variation in γ needs a complete response of capital accumulation to changes in γ to be able to generate a coefficient on capital equal to 1.

In addition, the implied differences in the rate of growth of technological change across countries seem implausibly high. For example, the growth rate of capital in Japan from 1950 to 1979 was 8.3 percent per year, in England during the same interval, 3.2 percent per year. If $\alpha = 1/3$, this implies a difference in the exogenous rate of technological change of more than $(2/3)5$ percent = 3.3 percent per year over a twenty-nine-year period. Thus, if Japan had accumulated capital at the same rate as England, exogenous technological change would cause income in Japan to increase by a factor of roughly $5.4 = 1.06^{29}$, as compared to the increase of $2.4 = 1.03^{29}$ actually achieved in England. Or stated the other way, even if Britain had invested in physical capital at the same rate as the Japanese, the neoclassical model implies that Britain would still have grown at a rate that was slower by 3 percent per year.

Figure 9 carries the international comparison of growth rates to the broadest possible context, offering comparisons of per capita income for 115 countries with market economies using the data from Summers and Heston (1984). Data for the full sample are available only from 1960 on. If the cross-sectional version of the Solow model is correct, diminishing marginal productivity should cause countries with low per capita capital to grow faster, converging to the level of the worldwide leader. Figure 9 plots a comparison of growth rates after 1960 against the ratio of per capita income in 1960 to per capita income in the United States. If convergence were a widespread phenomenon, this plot should show evidence of a downward slope. No such evidence is present.

The neoclassical theory predicts that low capital countries should catch up because (1) higher interest rates should induce higher domestic savings, (2) higher rates should attract foreign investment, and (3) the marginal productivity of a unit of invested capital is higher. Variation in government policy could easily thwart either of the first two effects from

operating, but it is possible to offer an indirect test for the third effect that is independent of the other two.

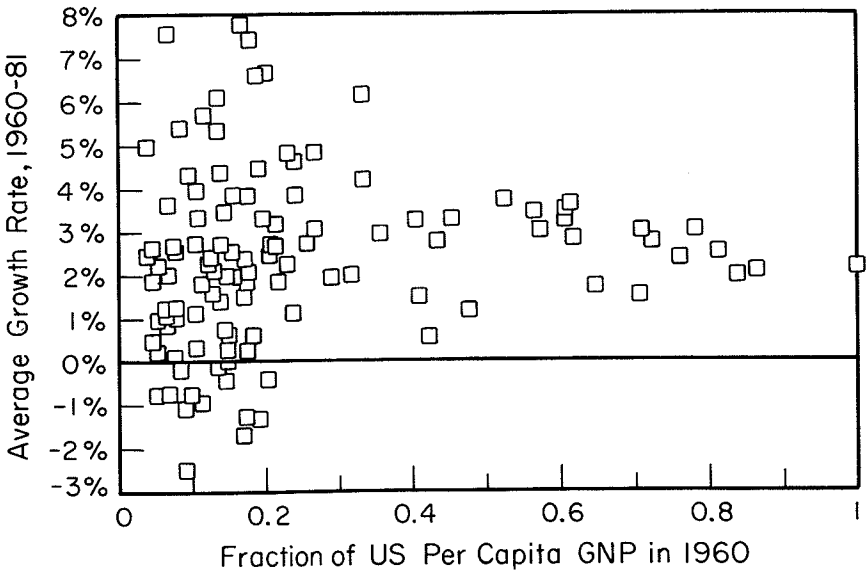
The Summers and Heston data include estimates of population and the share of GNP devoted to investment in each year, but not any direct capital stock measures. If output takes the form $Y(t) = \Gamma(t)K(t)^\alpha L(t)^{1-\alpha}$ and we use the identifying assumption that the rate of change of technology is the same in all countries, the correct way to estimate the parameter α is to regress \hat{Y} on a constant, \hat{K} , and \hat{L} . Since the gross rate of investment as a fraction of total output is observed, but the stock of capital is not, this cannot be done directly. If there were separate observations on the capital-output ratio for each country, this would be sufficient to infer \hat{K} . These data are not available, but provided that one is willing to assume that the level of the technology $\Gamma(t)$ is the same in all countries, an indirect estimate of Y/K is possible.

The ratio Y/K is given as a function of the capital-labor ratio,

$$\frac{Y}{K} = \Gamma \left[\frac{K}{L} \right]^{\alpha-1}. \quad (19)$$

Per capita output is also a function of this ratio,

Figure 9 115 COUNTRIES, SUMMERS AND HESTON



$$\frac{Y}{L} = \Gamma \left[\frac{K}{L} \right]^\alpha. \quad (20)$$

Solving for the capital-output ratio as a function of output per capita, yields

$$\frac{Y}{K} = \Gamma^{1/\alpha} \left[\frac{Y}{L} \right]^{\frac{\alpha-1}{\alpha}}. \quad (21)$$

Letting i denote the ratio of gross investment to output, and letting δ denote the rate of depreciation, the derived expression for \hat{K} is

$$\hat{K} = i \Gamma^{1/\alpha} \left[\frac{Y}{L} \right]^{\frac{\alpha-1}{\alpha}} - \delta. \quad (22)$$

Substituting this into the usual growth accounting equation leads to the nonlinear equation

$$\hat{Y} = \hat{\Gamma} - \alpha\delta + (1 - \alpha)\hat{L} + i\alpha\Gamma^{-1/\alpha} \left[\frac{Y}{L} \right]^{\frac{\alpha-1}{\alpha}}. \quad (23)$$

One should not have much faith in the reliability of this equation. Its only defense is that it presumably is an improvement over the more common regression of \hat{Y} on i (and possibly other variables) because it allows for some possibility that diminishing returns causes investment i to have a larger effect in a country with low per capita income (and hence presumably low per capita capital).

The growth rates Y and L were calculated for the interval for which all countries had data, 1960 to 1981. To minimize simultaneity problems, the values for Y and L were taken to be the values in 1960, denoted here as Y_{60} and L_{60} .³⁰ The investment fraction $i = I/Y$ was taken to be the arithmetic average of the values for i for the twenty-one years of data.

The equation actually estimated is

$$\hat{Y} = a + (1 - \alpha_1)\hat{L} + bi \left[\frac{Y_{60}}{L_{60}} \right]^{\frac{\alpha_2-1}{\alpha_2}} \cdot L_{60}^c + \varepsilon. \quad (24)$$

30. Measurement error in the level of Y_{60} or L_{60} would tend to introduce a spurious negative correlation between growth rates and the initial level of per capita income. Since this negative relation is not observed in the data, this cannot be the major source of variation across countries.

The parameters α_1 and α_2 were estimated separately to test the restriction that they are in fact equal. Similarly, the term L_{60} raised to the exponent c was added as a test of the restriction that the exponent on L_{60} is the negative of the exponent on Y_{60} . The disturbance term was assumed to be independent and normally distributed with standard deviation σ , and the equation was estimated using maximum likelihood. Under the assumptions of the neoclassical model, the restrictions are that the coefficients α_1 and α_2 should be the same and the exponent c on L_{60} should be zero. The estimate of α should be on the order of $1/4$ to $1/3$.

The results of constrained and unconstrained estimates are reported in table 4. As above, the units of \hat{Y} and \hat{L} are annual average percentage rates of growth times 100. Since Y_{60} and L_{60} were measured as indices, there is no obvious interpretation for the coefficient b . In the table, missing parameter estimates indicate that the relevant restriction has been imposed.

For an equation based on such crude assumptions and applied to the set of all market economies in the world, the results are surprisingly reasonable and fairly consistent with the previous results. In particular, the capital elasticity is precisely estimated and is much larger than $1/3$.

Since each country is given equal weight in this sample, it is weighted heavily toward countries with very low levels of per capita capital and income. (The mean of the ratio of per capita income in each country to that in the United States is around 0.25.) The evidence from these countries suggests the following qualifications for the conclusions drawn from developed countries. First, considered over values of the per capita capital ranging from those for Bangladesh to those for the United States, the elasticity of output with respect to capital appears to be closer to 0.75 than 1. Thus, there is some evidence of diminishing returns, that the

Table 4 COEFFICIENT ESTIMATES FOR 115 MARKET ECONOMIES, 1960-1981

a	α_1	b	α_2	c	σ	Log likelihood
-.27 (.74)	.51 (.23)	12.0 (2.5)	.77 (.04)	.01 (.02)	1.5	538.6
-.24 (.74)	.51 (.23)	11.5 (1.9)	.77 (.04)		1.5	538.5
.35 (.47)	.74 (.03)	10.2 (1.8)			1.5	537.7

Note: Standard errors in parentheses. See text for equation.

marginal product of capital may indeed be higher in the least developed countries. However, even in this sample, the diminishing returns are not nearly as large as the conventional estimates suggest. Second, there is weak evidence that the elasticity of output with respect to labor may be larger than previous estimates suggest. The unrestricted estimate is on the order of 0.5, but it is not precisely estimated.

4. *Conclusions*

As outlined in the introduction, the tentative conclusion I draw from this exercise is that the appropriate growth accounting equation is $\hat{Y} = \alpha \hat{K} + \beta \hat{L}$, with values for α likely to fall in the range 0.7 to 1.0, and values for β likely to fall in the range 0.1 to 0.3, with values possibly as large as 0.5. For developed countries, higher values for α and lower values for β may be more likely.

For the explanation of the productivity puzzle, the key implication of this revised interpretation of growth accounting is that an increase in the rate of growth of labor will be accompanied by a fall in the rate of growth of labor productivity. This may explain the productivity slowdown in the United States since the 1960s. It may also help explain international comparisons of productivity growth rates, although in cases such as Japan during the 1960s, differences in capital accumulation must also have played a role.

The policy implications of this view are too numerous to be detailed here and deserve a separate treatment. But since policy is ultimately what is at stake, it may be useful to indicate how radically these alternatives may challenge conventional presumptions. For example, there may be a rationale behind the policies followed in Europe to reduce total hours worked and to keep wage growth rates high, although it has had undeniable social and distributional costs. Perhaps we should take more seriously the European view noted in Mairesse (1982) that there is a trade-off between productivity growth and unemployment.

Or, if the divergence between the social and private rates of return on capital is anywhere near as large as suggested here, what matters most for national wealth is investment, not savings. Holding constant the level of consumption (and this caveat is essential) the advantages of being a net foreign debtor are enormous; we should be running as deep a trade deficit as possible for as long as we can get away with it. According to Maddison, by the time that each was surpassed as a world leader in productivity, the Netherlands had accumulated net foreign assets of over 3 times GDP, the U.K. assets totaling roughly 1.5 times GDP. Had the United States accumulated net foreign assets of this magnitude, divert-

ing domestic investment abroad, it too may have been surpassed by a new borrowing country. So perhaps, because of recent economic policies, the United States really is back, standing tall, rapidly accumulating foreign debt.

REFERENCES

- Allen, S. G. 1985. Why construction industry productivity is declining. *Review of Economics and Statistics* 67 (November): 661–69.
- Arrow, K. J. 1962. The economic implications of learning by doing. *Review of Economic Studies* 29 (June): 155–73.
- Baily, M. N. 1981. Productivity and the services of capital and labor. *Brookings Papers on Economic Activity* 1:1–50.
- . 1986. Productivity and materials use in U.S. manufacturing. *Quarterly Journal of Economics* 101 (February): 185–95.
- Baily, M. N. and A. K. Chakrabarti. 1985. Innovation and productivity in U.S. industry. *Brookings Papers on Economic Activity* 2:609–33.
- Bernstein, J. I. and M. Ishaq Nadiri. 1986. Research and development and intra-industry spillovers: An empirical application of dynamic duality. National Bureau of Economic Research Working Paper no. 2002, August.
- Bruno, M. 1984. Raw materials, profits, and the productivity slowdown. *Quarterly Journal of Economics* 99 (February): 1–12.
- Darby, M. R. 1984. The U.S. productivity slowdown: A case of statistical myopia. *American Economic Review* 74 (June): 301–22.
- De Long, J. B. 1987. Have productivity levels converged? M.I.T. (March). Mimeo.
- Dixit, A. K. and J. Stiglitz. 1977. Monopolistic competition and optimum product diversity. *American Economic Review* 67 (June): 297–308.
- Drandakis, E. M. and E. S. Phelps. 1966. A model of induced invention, growth and distribution. *Economic Journal* 76 (December): 832–40.
- Ethier, W. J. 1982. National and international returns to scale in the modern theory of international trade. *American Economic Review* 72 (June): 389–405.
- Fellner, W. J. 1961. Two propositions in the theory of induced innovations. *Economic Journal* 81 (June): 305–8.
- Friedman, Milton and A. J. Schwartz. 1982. *Monetary trends in the United States and the United Kingdom*. Chicago: NBER, University of Chicago Press.
- Gallman, R. E. 1966. Gross national production in the United States, 1834–1909. In *Output, Employment, and Productivity in the United States after 1800*. NBER Studies in Income and Wealth, vol. 30. New York: Columbia University Press.
- Gollop, F. M. and D. Jorgenson. 1980. U.S. productivity growth by industry, 1947–73. In *New Developments in Productivity Measurement and Analysis*, ed. John W. Kendrick and Beatrice N. Vaccara. NBER Studies in Income and Wealth, vol. 44. Chicago: University of Chicago Press.
- Gordon, R. J. 1983. The productivity slowdown in the steam-electric generating industry. Northwestern University Working Paper (February).
- . 1985. Wage-price dynamics and the manufacturing output gap in Europe, Japan, and North America. Northwestern University Working Paper (September).
- . 1986a. Productivity, wages and prices inside and outside of manufacturing in the U.S., Japan, and Europe. National Bureau of Economic Research Working Paper no. 2070, November.

- Gordon, R. J. 1986b. Measurement of durable goods prices. Northwestern University. Ms.
- Griliches, Z. 1980. R & D and the productivity slowdown. *American Economic Review* 70 (May): 343–48.
- . 1986. Productivity, R & D, and basic research at the firm level in the 1970s. *American Economic Review* 76 (March): 141–54.
- Habakkuk, H. J. 1962. *American and British technology in the nineteenth century*. Cambridge: Cambridge University Press.
- Hall, R. E. 1986. Productivity and the business cycle. Hoover Institute Working Paper, November.
- Hicks, J. 1963. *Theory of Wages*. London: Macmillan.
- Hollander, M. and D. A. Wolfe. 1973. *Nonparametric statistical methods*. New York: Wiley.
- James, J. A. and J. S. Skinner. 1985. "The resolution of the labor-scarcity paradox," *Journal of Economic History* 45 (Sept.): 523–41.
- Jorgenson, D. W., F. M. Gollop and B. M. Fraumeni. 1986. Productivity and sectoral output growth in the United States. In *Interindustry differences in productivity growth*, ed. John W. Kendrick. Cambridge: Ballinger.
- Kendrick, J. W. 1961. *Productivity trends in the United States*. Princeton: Princeton University Press.
- Kennedy, C. 1964. Induced bias in innovation and the theory of distribution. *Economic Journal* 84 (Sept.): 541–47.
- King, R. G. and S. Rebelo. 1986. Business cycles with endogenous growth. University of Rochester. Mimeo.
- Kuznets, S. 1971. Notes on the pattern of U.S. economic growth. Reprinted in *The Reinterpretation of American Economic History*, ed. Robert W. Fogel and Stanley L. Engerman. New York: Harper and Row.
- Lebergott, S. 1966. Labor force and employment, 1800–1960. In *Output, Employment, and Productivity in the United States after 1800*, NBER Studies in Income and Wealth, vol. 30. New York: Columbia University Press.
- Lichtenberg, F. R. and Z. Griliches. 1986. Errors of measurement in output deflators. National Bureau of Economic Research Working Paper no. 2000, (August).
- Lucas, R. E., Jr. 1985. On the mechanics of economic development. Prepared for the Marshall lecture (May).
- Maddison, A. 1982. *Phases of capitalist development*. Oxford: Oxford University Press.
- Mairesse, J. 1982. Comments on "Economic policy in the face of declining productivity growth" by Nordhaus. *European Economic Review* 18 (May): 159–62.
- Mark, J. A. 1986. Problems encountered in measuring single- and multifactor productivity. *Bureau of Labor Statistics Monthly Labor Review* (December): 3–11.
- Mark, J. A. and W. H. Waldorf. 1983. Multifactor productivity: A new BLS measure. *Bureau of Labor Statistics Monthly Labor Review* (December): 3–15.
- Nordhaus, W. D. 1982. Economic policy in the face of declining productivity growth. *European Economic Review* 18 (May): 131–57.
- Norsworthy, J. R. 1984. Growth accounting and productivity measurement. *Review of Income and Wealth*: 309–29.
- Oshima, H. T. 1984. The growth of U.S. factor productivity: The significance of new technologies in the early decades of the twentieth century. *Journal of Economic History* 44 (March): 161–70.

- Phelps, E. S. 1966. Models of technical progress and the golden rule of research. *Review of Economic Studies* 33 (April): 133–45.
- Prescott, E. 1986. Theory ahead of business cycle measurement. Federal Reserve Bank of Minneapolis, Research Department Staff Report 102 (February).
- Rebelo, S. 1987. Long-run policy analysis and long-run growth. University of Rochester (March).
- Romer, P. M. 1986a. Increasing returns and long-run growth. *Journal of Political Economy* 94 (October): 1002–1037.
- . 1986b. Increasing returns, specialization, and external economies: Growth as described by Allyn Young. Rochester Center for Economic Research Working Paper no. 64.
- . 1987. Growth based on increasing returns due to specialization. *American Economic Review*, forthcoming, May 1987.
- Romer, P. M. and H. Sasaki. 1986. Scarcity and growth reinterpreted: Endogenous technological change and falling resource prices. Rochester Center for Economic Research Working Paper no. 19, revised October 1986.
- Rosenberg, N. 1976. *Perspectives on technology*. Cambridge: Cambridge University Press.
- Rothbart, E. 1946. Causes of the superior efficiency of U.S.A. industry as compared with British industry. *Economic Journal* 56 (Sept.): 383–90.
- Samuelson, P. A. 1965. A theory of induced innovation along Kennedy-von Weizsacker lines. *Review of Economics and Statistics* 47 (November): 343–56.
- Sato, K. 1966. On the adjustment time in neo-classical growth models, *Review of Economic Studies* 33 (July): 263–68.
- Sato, R. 1963. Fiscal policy in a neo-classical model: An analysis of time required for equilibrating adjustment, *Review of Economic Studies* 30 (Feb.): 16–23.
- Schmookler, J. 1966. *Invention and economic growth*. Cambridge: Harvard University Press.
- Shell, K. 1967. A model of inventive activity and capital accumulation. In *Essays in the Theory of Optimal Economic Growth*, ed. Karl Shell. Cambridge: MIT Press.
- Solow, R. 1957. Technical Change and the Aggregate Production Function. *Review of Economic Studies* 39 (August): 312–20.
- Solow, R. 1959. Investment and technical progress. In *Mathematical Methods in Social Sciences*, ed. Kenneth J. Arrow, Samuel Karbin, and Patrick Suppes. Palo Alto: Stanford University Press.
- Summers, R. and A. Heston. 1984. Improved international comparisons of real product and its composition: 1950–1980. *Review of Income and Wealth* (June): 207–62.
- Sveikauskas, L. 1986. The contribution of R & D to productivity growth. *Monthly Labor Review* 109 (March): 16–20.
- Temin, P. 1966. Labor scarcity and the problem of American industrial efficiency in the 1850s. *Journal of Economic History* 26 (Sept.): 227–98.
- Waldorf, W. H., K. Kunze, L. S. Rosenblum, and M. B. Tannen. 1986. New measures of the contribution of education and experience to U.S. productivity growth. Bureau of Labor Statistics Working Paper.
- Wallis, J. J. and D. C. North. 1986. Measuring transaction costs in the American economy, 1870–1970. In *Long term factors in American economic growth*, ed. Stanley L. Engerman and Robert G. Gallman. NBER Studies in Income and Wealth, vol. 51. Chicago: University of Chicago Press.

Weiss, T. 1986. Revised estimates of the United States workforce, 1800–1860. In *Long term factors in American economic growth*, ed. Stanley L. Engerman and Robert G. Gallman. NBER Studies in Income and Wealth, vol. 51, Chicago: University of Chicago Press.

von Weizsacker, C. C. 1966. Tentative notes on a two sector model with induced technological progress. *Review of Economic Studies* 33 (July): 245–51.

Comment

BEN S. BERNANKE

Woodrow Wilson School, Princeton University

Paul Romer's stimulating paper tackles the most important subject in economics: the determinants of long-run growth. Not surprisingly, given his recent work, Romer studies the problem in the framework of optimal growth models with increasing returns. After presenting a useful survey of these models, he informally considers their consistency with a broad range of evidence. His major conclusions are, first, that it may be worthwhile to dispense with the notion of exogenous technical change and instead attribute all growth of average productivity to increasing returns; and, second, that an important part of the growth process may be labor-saving technical changes induced by shortages of labor. The second conclusion has the implication that there may be a trade-off between high employment and productivity growth.

I will first take up some issues related to increasing returns in general, then turn to Romer's empirical evidence. Finally, I will comment on an element missing from the present article that may be essential to a comprehensive analysis of growth.

A number of increasing returns stories are told in the article, all of which contain significant elements of truth. A distinction that should be made, though, is the one between increasing returns over a range (which everyone accepts) and universal increasing returns. For example, given human mortality and limits on processing capability, is it meaningful to assume that human capital per worker can grow without bound (as in the models of Uzawa and Lucas)? Similarly, Romer's horseshoe example notwithstanding, most significant innovations in given industries occur early in the industry life cycle; innovative energies then pass on to less-worked fields. This extremely well-documented industry life-cycle pattern (Norton 1986) can easily explain the cited results of Schmookler that the rates of patenting in technologically unrelated parts of a given industry are correlated, and that patenting and investment generally are correlated. Romer claims that this life-cycle pattern is demand-driven, but the burden of proof is on him to show that none of this pattern is related to

the exhaustion of technological opportunities. (On the other hand, Romer does not need to claim necessarily that opportunities within a given industry and technology are never exhausted, only that there will always be new industries and new technologies to replace the old ones. The experience of the last century contains nothing to controvert this broader view of increasing returns.)

The increasing returns story based on ever-growing specialization is appealing. I would think that the cited estimate that the proportion of the economy devoted to transactions (broadly defined) has grown from 25 percent in 1870 to 45 percent in 1970 is understated; it is well known that no one in the United States actually ever makes anything. This is also consistent with Raymond Goldsmith's frequently cited findings that the relative size of the financial sector and the share of wealth devoted to transactions balances increases with economic development.

Whether increasing returns are pervasive or not makes a tremendous difference for economic analysis and policy. Consider North-South relations: As Lucas has recently pointed out, with increasing returns, the late entry of most LDCs into the economic race may prevent them from ever competing with the developed nations. They will instead just fall further and further behind—unless, like the NICs, they can make the jump into increasing returns, export-oriented industries. (With increasing returns, one can no longer dismiss the claim that colonial exploitation decades, or even centuries, ago is responsible for such-and-such a country's current abject condition.) Nor are the LDCs the only ones who need worry: American business leaders have expressed the concern that, because of long corporate planning horizons and a putatively greater ability to sustain short-run losses, the Japanese will soon be able to achieve dominance in certain computer technologies for which learning-by-doing is essential. (Paul Krugman has been studying this issue.) This could lead to a perpetual inferiority of U.S. producers in a range of products that could ultimately be very profitable. The policy recommendations from increasing returns include aggressive development and "industrial policy" type strategies. With increasing returns, being the leader is important; without them, it is easier to follow.

Thus we would very much like to know the empirical relevance of the increasing returns phenomenon. Unfortunately, the evidence presented here is not conclusive. A major problem is the reliability and comparability (for example, across countries) of the data. It would be useful, for example, to think a bit about the meaning of those artificial constructs, "output," "capital," and "labor," when they are measured over such long time periods (the Cambridge-Cambridge debate and all that). What does it mean to say that the capital stock was (say) twenty times larger in 1980

than in 1880? How many spinning jennies equal a personal computer?

Perhaps the principal empirical finding of the article is that, in low-frequency data, the exponent on the capital stock in the Cobb-Douglas production function is close to one, while the exponent on labor is close to zero—almost exactly the opposite of what one finds in high-frequency data. Since it cannot literally be true that output is independent of labor input, this result must be caused by an estimation bias; presumably, capital inputs are positively correlated, and labor inputs negatively correlated, with the regression equation's residual.

Now this residual is of course the same as Denison's famous residual: it captures all the influences on measured output that are not included as explicitly measured inputs, including the state of technology, unmeasured input quality, aggregation errors, and so on. Romer's interpretation of the apparent correlations of capital and labor with this residual is that changes in capital and labor cause changes in the residual, either because the investment process has spillovers, or because labor shortages induce labor-saving technical change, or both. The alternative is, of course, that the residual is at least partly due to forces exogenous to the levels of measured inputs; and that variations in the residual induce behavioral responses in K and L . For example, a positive productivity shock might induce greater investment (because capital is more productive) and less labor input (through the wealth effect).

A way to get around this identification problem is to look for exogenous changes in inputs. Wars, for example, are exogenous events that sometimes lead to large changes in national capital stocks or labor force composition (for example, through baby-boom effects; see Bloom and Freeman 1986). A cross-national study of the effects of wars on subsequent economic growth might be revealing. Another possibility would be to see if there is a positive correlation in the residual across countries when measurable inputs are held constant.

The empirical work treats business cycles as exogenous; Romer even uses cyclical fluctuations as an "explanation" of the recent slowdown in growth. Romer probably knows that he has colleagues at Rochester who treat growth shocks as exogenous and use them to explain cycles. Perhaps a joint paper is in order. Romer is closer to his colleagues in recognizing that increasing-returns models can generate unit roots in output. The apparent importance of unit roots in macro time series is evidence in favor of the increasing returns approach.

Finally, let me explain why there may yet be a missing piece to the growth puzzle. As William Baumol has emphasized, the most striking comparison of all is between the growth rates of industrialized countries during the last century and the much slower growth of previous millen-

nia. (Ancient Roman standards of living were close to those of eighteenth-century England, according to Baumol.) Why did the growth process erupt at the time that it did? Why did it take a firmer hold in some places than in others? Why has economic leadership changed from country to country and region to region? Why have some countries (such as Argentina) appeared to regress toward underdevelopment?

I doubt that the austere and ahistorical "production function approach" of this article (which states that factor productivity is a function of the level of measured inputs and is independent of time or place) can go very far toward answering these questions. (I could be proven wrong.) More likely, we economists may have to face the fact that social and political evolution is an integral part of economic development. This does not mean that we all have to become sociologists (though we should learn from them); a good part of the social and political change accompanying growth is endogenous in the usual sense and could be modeled as such. A theory of societal stability may be an essential part of a theory of economic growth. One immediate implication of such a theory might be an alternative to Lucas's human capital externalities explanation of why both labor and capital move toward developed countries. That such a theory could explain short-run fluctuations, like the current slowdown, is possible but less likely.

REFERENCES

- Bloom, D. and R. Freeman. 1986. The 'youth problem': Age or generational crowding? HIER discussion paper 1223, April.
 Norton, R. D. 1986. Industrial policy and American renewal. *Journal of Economic Literature* 24:1 (March), 1-40.

Comment

MARTIN NEIL BAILY
 The Brookings Institution

I like the new classical growth theory. Paul Romer and Robert Lucas are asking fundamental questions about what it means to analyze a national economy as a separate entity. They pose two puzzles that they say are inconsistent with the way in which traditional neoclassical growth theory treats national economies. The first is that workers migrate from LDCs to industrialized countries, whereas capital does not flow strongly toward LDCs. Most capital flows are among advanced economies. The second puzzle has to do with the convergence of economies.

A worker coming to the United States from a developing country may

triple her salary or more. Rates of profit, on the other hand, have shown no such disparity. Indeed, the United States has traditionally had rates of return that were as high or higher than those in countries with much smaller capital-labor ratios. Recently, capital has actually been flowing into the United States.

This is an old issue, of course. The classic cross-country study by Arrow, Chenery, Minhas, and Solow had trouble with this point. But the issue is a good one and needs to be confronted. The explanation offered by Paul Romer is that there is an externality associated with the physical capital stock. The capital held by any single company receives a direct benefit from the existence of capital elsewhere in the economy. In Romer's preferred model for the U.S. economy, the coefficient on capital in a production function is unity. This means that the positive effect of the externality just offsets diminishing returns. There is no tendency for rates of return to be lower in high-capital economies.

Having been taught by Joan Robinson and Nicholas Kaldor in my student days, I am intrigued to see this proportionality assumption emerge from Chicago and Rochester. Joan Robinson was adamant about the proportionality of output and capital (or at least the proportionality of investment and the increment to output). I remain unconvinced, however.

The basic problem is that Romer's model does not fit the facts of the postwar U.S. economy. Romer proposes that an appropriate growth accounting framework is $\hat{Y} = \alpha\hat{K} + \beta\hat{L}$, where α is between 0.7 and 1.0 and β is between 0.3 and 0.1. He prefers 1.0 and 0.1 for the United States (the notation is his). Data from the Bureau of Labor Statistics (BLS) for the nonfarm business economy give the following results from his specification:

<i>Period</i>	<i>Actual \hat{Y}</i>	<i>Estimated \hat{Y} $\alpha=1.0 \beta=0.1$</i>	<i>Residual</i>	<i>Estimated \hat{Y} $\alpha=0.7 \beta=0.3$</i>	<i>Residual</i>
1948-65	3.72	3.05	0.67	2.37	1.35
1965-73	3.65	4.62	-0.98	3.65	-0.01
1973-85	2.31	3.86	-1.56	3.08	-0.77

Clearly his parameter values do not track postwar output growth. They generate the same large unexplained residuals that exist in conventional growth accounting models. Indeed the slowdown is even greater in Romer's residuals than in the BLS's conventional estimate of multifactor productivity growth.

The second problem with proportionality is that the nature of the externality inherent in physical capital is never made plausible. Romer cites Schmookler's work to the effect that high rates of patenting and investment are correlated. But this does not show an externality. Even if we

accept the Schmookler view, the explanation is that a third factor, namely demand, is driving both patents and investment. The critics of Schmookler, and there are many, suggest that the correlation arises because both are driven by some major initiating innovation.

There surely is a major externality involved in the knowledge factor of production, but the literature has usually and, I think, correctly associated this with R & D and innovation. This is a literature that Romer leaves untouched.

The empirical evidence Romer offers to support proportionality is in two parts. He uses time-series analysis for the United States and gets some results consistent with it. The support is pretty weak here, however, as minor changes in specification cause the coefficients to jump. All of us who have estimated time-series production functions can sympathize with Romer in trying to separate trend and cycle and estimate reliable coefficients. It is not clear to me that time-series data can identify the relevant parameters. In particular, steady-state growth in the Solow model also implies the proportionality of output and capital.

Romer next looks at data on over 100 countries. And this analysis leads to the second of the two original objections that were made of conventional growth analysis. Within this broad sample of countries, there is no tendency for convergence, contrary to the prediction of the Solow model. Instead, Romer finds a strong tendency for output growth to move one-for-one with capital growth. Low-income, low-capital economies are not catching up.

This criticism flies in the face of a conventional wisdom that indeed convergence has taken place. Europe and Japan have largely caught up. Romer points out, however, that convergence is a phenomenon of only this small group of countries.

I am not sure how compelling an objection this is to orthodox theory. If you had told Bob Solow back in 1956 that his theory would work for the United States, Europe, and Japan, but miss for Swaziland, he would have been pretty happy. In fact, the natural response is to look for special factors applicable to the LDCs, and the theory of dual economies with surplus labor has been proposed as just such a variation on conventional theory. In labor surplus economies, the proportionality of output and capital is plausible. Joan Robinson's model was essentially one with surplus labor.

So if I do not believe Romer's explanation of why labor migrates but capital does not, what is the explanation? I am not sure I know, but the idea of externalities is certainly correct. The advantage of being an American worker is tied to the body of technical knowledge and organizational knowledge that exists here and does not exist in other countries. Con-

trary to Romer's assertion, this knowledge is not available worldwide, and even if it was, it could not be applied without the human capital and experience possessed in developed but not in underdeveloped countries.

Finally, I want to issue a word of warning before anyone accepts the hypothesis that U.S. manufacturing now shows no persistent productivity slowdown. First, the slowdown in multifactor productivity growth in manufacturing is greater than that in labor productivity. Second, the recent recovery in manufacturing is very heavily tied to a single industry. Taking nonelectrical machinery (the 2-digit industry with computers in it) out of manufacturing drops the rate of growth in multifactor productivity for the resulting aggregate by a full percentage point, 1979–1985. Fifteen out of twenty 2-digit manufacturing industries had substantially slower growth 1979–1985 than the rate they achieved over the entire period 1948–1985. Third, the pressure of foreign competition created a Darwinian environment that closed many inefficient plants and boosted productivity in the period 1981–1985. History may show that manufacturing had no slowdown. But let's wait a while to be sure.

In his choice of title for this paper and in various disclaimers scattered through it, Paul Romer indicates his own skepticism about the detailed hypotheses he advances. He is sending up trial balloons and asking us to think in new ways about growth. He succeeds admirably in this task even though my own skepticism about the specifics is greater than his.

Discussion

Zvi Griliches stressed the difficulty of obtaining empirical evidence that supports growth theories based on increasing returns or externalities. In the short run, the coefficient on capital is small and far from one. In the long run, various simultaneity problems make estimation difficult. These externality or spillover stories must be important in the long run, but they do not explain short-run phenomena like the productivity slowdown.

He also commented on Schmookler's evidence. His own work on investment, patents, and R&D shows that we could reproduce Schmookler's results. However, the timing suggested by externality theory is not supported by the data. What is important is the impact of expectations about the fortunes of an industry, which drives both investment and technology more or less at the same time. These expectations tend to be created either by changes in demand for products or by a technological breakthrough. It is hard to figure out causality in these simultaneous events.

Lawrence Summers commented that Romer's model should apply not only to the international economy but also to the different regions of the United States economy. He mentioned two impressions that are somewhat contradictory to the implications of Romer's model. First, there has been a convergence in income levels among different regions in the United States. Second, the growth of productivity does not differ much among regions, although growth rates of capital differ a lot.

Robert Gordon commented that Romer's explanation of the productivity slowdown does not work for Europe and Japan. Both had a much greater productivity slowdown than the United States, yet they experienced no acceleration in employment. Romer's econometric result that the coefficient on capital is one implies that a country can increase growth simply by investing more, which is not the case. Finally, he questioned the advantage of being a leader that is implied by the increasing returns story. If we had compared Argentina and Korea in 1979, the increasing returns theory would have predicted that Argentina would move progressively farther ahead.

Susan Collins suggested looking at some countries that are not experiencing a productivity slowdown. An example is South Korea, where there is a strong relation between investment and growth.

Stanley Fischer mentioned an alternative view of convergence, citing Milton Friedman. With immigration to higher-income countries, per capita income of countries might not converge, even if people's incomes were converging. He also commented that the correlation between the growth rate and the investment share may be special to developed countries. He had not been able to find the relationship in a cross-sectional regression for developing countries over the past twenty years.

Patrick Minford proposed an explanation for migration to richer countries without assuming increasing returns. Although the wage a firm pays to a worker is equal to the marginal product of labor, the total benefit a worker receives in a richer country can be much larger because of the superior public goods available there. This can lead to migration, which is desired by workers in the poorer country and undesired by people in the richer country.

Answering some of the comments, Paul Romer first stressed that the increasing returns story is not the essential point of his article, which relates more to the role of capital. If capital growth is incorporated in the regression in the right way, we can detect a strong relationship between investment and growth. Answering Gordon's comment, he suggested that the Japanese productivity slowdown can be explained by the change in the rate of growth of capital in Japan. Explaining the productivity

slowdown in Europe is harder. He also stressed the difficulty of sorting out cyclical movements from long-run growth. He agreed with Bernanke on the importance of looking at the effect of exogenous changes in capital growth or in labor supply on growth, though the suggestion is very hard to implement.