PAUL M. ROMER
University of Chicago
and Center for Advanced Study in the Behavioral Sciences

Capital, Labor, and Productivity

IN A PREVIOUS PAPER I explored two suggestions about how to understand time-series and cross-country variations in measured total factor productivity growth: increases in the labor force might slow technological change and increases in capital might speed it up.[1] Neither suggestion was new. The conjecture about the effect of labor dates back, at least, to attempts to explain the divergence in productivity growth rates observed in the United States and the United Kingdom.[2] The suggestion that investment or savings is a fundamental determinant of the rate of growth dates back to Adam Smith. Neither possibility can be considered within the narrow theoretical confines of neoclassical growth theory, but more recent models of endogenous growth show that they can arise in richer economic environments.

This paper presents new evidence and new theoretical arguments that bear on these matters. In the theoretical model presented here, the rate of technological change depends on the amount of educated human capital devoted to applied research and development, which is interpreted in a broad sense. The model confirms the conjecture that an increase in the labor force can reduce the rate of technological change under appropriate assumptions about the possibilities for substitution between capital goods, physical labor, and skilled human capital, for example in the form of managers. The rate of improvement in the

A discussion of the basic model with Kiminori Matsuyama was very helpful, as were comments from Martin Baily, Luis Rivera, and Danyang Xie. This work was supported by a Sloan Foundation Fellowship, and National Science Foundation grants to Paul Romer and to the Center for Advanced Study in the Behavioral Sciences.

1. Romer (1987b).
2. Rothbart (1946); and Habakkuk (1962).

technology falls if the elasticity of substitution between physical capital and a composite of workers and managers is higher than the elasticity of substitution between workers and managers. For example, this condition will hold if workers and managers are used in fixed proportions and if capital can be used to substitute for both types of human inputs.

The model also shows that an increase in the rate of accumulation of physical capital may have no effect on the rate of technological change. This contrasts with results from previous models of endogenous technological change that rely on learning by doing in capital accumulation or on the idea that technological innovation and physical capital are such strong complements that an increase in the rate of growth of physical capital necessarily leads to an increase in the rate of technological change.[3] In those models, a permanent increase in the share of GDP devoted to investment causes a permanent increase in the rate of growth of technological change and of output. In the model presented here, an increase in the investment share has no long-run effect on the growth rate of the technology or output.

In making the claim that faster capital accumulation could lead to faster growth, I drew some theoretical support from the early endogenous growth models.[4] The main support, however, came from the data, specifically from the positive correlation across countries between the investment share and the growth rate. This cross-sectional correlation is one of the most robust correlations (or partial correlations) to emerge from an analysis of large cross-country data sets. It obtains in data for the past 100 years for developed countries. It is also evident in postwar data for a large sample of developing countries. It does not appear to be a purely transitory effect; it remains evident in data that are averaged over 25- to 40-year intervals.

This correlation has at least two possible interpretations. When I wrote the previous paper, my prior belief was that much of the variation in savings rates across countries is not caused by cross-country variation in the rate of technological change. If most of the variation in savings rates is exogenous in this sense, the cross-country correlation leads to a presumption of causality that runs from savings and investment to

3. Romer (1986).
4. Romer (1986).

growth. The evidence that follows runs counter to this interpretation.

The empirical analysis presented here looks in more detail at the cross-country correlation between investment and growth. If increases in the rate of investment do not cause one-for-one increases in the rate of technological change, then they should cause a reduction in the marginal product of capital, just as they do in the neoclassical model with exogenous technological change. Consistent with this view, the regression evidence shows that increased investment tends to be correlated with a lower marginal product of physical capital. The precision of these estimates is not sufficient to conclude that increased investment has no effect at all on technological change, but at a minimum the estimates show that increased investment does not seem to induce enough technological change to offset completely the diminishing returns associated with increased capital accumulation. Thus contrary to the suggestion of those who place exclusive reliance on capital accumulation to generate long-run growth (and contrary to my previous claims)[5] something else—something that does not vary one-for-one with the rate of investment in physical capital—must be decisive for long-term productivity growth.

This finding opens up the possibility of making sense of the negative correlation between labor force growth and productivity growth. If capital accumulation were the main driving force behind technological change, an increase in the labor force would tend to increase productivity growth because of its conventional positive effect on capital accumulation. (This result is explicitly derived by Ken Arrow, who shows the rate of growth of output per capita is increasing with the rate of growth of the population.)[6]

Specification of the Technology

This section presents the theoretical model and derives the effects that an increase in capital accumulation and an increase in the labor force have on productivity growth and output growth.

5. Romer (1987b).
6. Arrow (1962).

Production of Final Output

The model used here is an extension of one I have outlined elsewhere.[7] The fundamental concepts are the two attributes possessed in varying degrees by any economic good: rivalry and excludability. An input in production is rival if its use by one person or firm precludes its use by another. A good is excludable if the owner can preclude others from making use of it. Conventional economic goods are rival and excludable and are privately provided. Public goods are neither rival nor excludable and cannot be privately provided. The interesting feature of technological advance is that it is alleged to have the characteristics of a public good but is privately provided by firms that do research and development.

Nonrivalry is a fundamental attribute of what is called the technology. It is in this sense that the technology resembles public goods. A new technology—for example a new list of instructions for a chemical process or a design for a new good—can be used as many times and in as many different production processes as desired. In the neoclassical model with technological change, output for firm j was expressed as a function $Y = A \cdot F(K_j, L_j)$ or $Y = F(K_j, AL_j)$ of capital, K_j, and labor, L_j, employed by the firm.[8] The technology index, A, is not indexed by the firm that uses it. All firms can make use of all of A at the same time.

Initial attempts to account for growth attributed a very large fraction of growth in output to growth in A. Methodological refinements have substantially raised the estimate of how important growth in effective capital and effective labor has been in generating growth in total output.[9] Nevertheless, the growth-accounting work of the late 1950s, most prominently that of Robert Solow, had the lasting effect of convincing economists that technological change was real, important, and different from the accumulation of rival inputs such as capital and labor.[10] One advantage of the neoclassical formulation of aggregate production was that it emphasized in the simplest and most forceful way that if one treats the technology as an input, it must be nonrival and must lead to

7. Romer (1990).
8. See, for example, Solow (1956).
9. Jorgenson, Gollop, and Fraumeni (1987).
10. Solow (1957).

a nonconvexity. By appeal to a simple replication argument, it follows that the production function, F, should be homogeneous of degree 1. Consequently, output increases more than proportionally with increases in all three inputs—K, L, and A.

If A is treated as a good that is both nonrival and completely nonexcludable—that is, as a pure public good—growth in A could arise either exogenously or as the result of government funding for basic science.[11] Because A is nonexcludable, it cannot be compensated for in the market. Models of learning by doing assume that the technology is both nonrival and nonexcludable but assume that technological advances are privately provided as an unintended side effect of some other activity. Arrow assumed that technological advance is a side effect of investment in capital; Robert Lucas in effect assumed that technological advance arises as a side effect of education.[12] The deficiency of the learning-by-doing formulation is that it cannot explain why research and development activity is undertaken intentionally by private firms.

Zvi Griliches tried to explain private research and development at the level of the firm and industry by treating a part of the technology as a conventional good.[13] In my first attempt to model growth, I also used this approach.[14] We both assumed that production for a single firm, j, takes the form $y_j = AG(a_j, k_j, l_j)$, where a_j represents research results that are specific to firm j, k_j is physical capital, and l_j is labor. This formulation allows for incomplete excludability by assuming that the spillover term, A, of which every firm can take advantage, is a function of the investments by all firms in a_j. To generate an equilibrium with competition, this approach requires the assumption that the function G is at most homogeneous of degree 1 in a, k, and l, but this runs counter to the essentially nonrival character of technology. For any fixed stock of A and a_j, firm j should be able to double its output by doubling only its rival inputs, k and l (and land and any other rival inputs that are relevant). If a is doubled as well, output should more than double. When I made the assumption that the function analogous to $G(\cdot)$ was homogeneous of degree 1, it seemed a harmless short cut.

11. Shell (1966, 1967).
12. Arrow (1962); and Lucas (1988).
13. Griliches (1979).
14. Romer (1986).

But glossing over this issue misses the essence of the problem of how technological change is generated in decentralized markets.

My more recent model captures both aspects of improvements in the technology: they are nonrival goods, and they are provided for the most part by private firms.[15] The result is a model in which technology is treated as a nonrival input that is partially excludable. Therefore there may be knowledge or technology spillovers, but technological advance is not a pure public good. Innovators capture at least part of the social benefits of an improvement in the technology.

The model makes this precise by representing the technology through a stock of designs for producer durables that are used in production. Research is necessary to make a design for a new type of durable. Once the design is complete, the durable can be produced with a production function that is homogeneous of degree 1. The design is nonrival because it can be used to make as many copies of the good as desired. A firm that owns a design and sells a new durable charges a price for the good that is higher than the constant cost of producing the good. This is how the firm recoups the investment in the research necessary to create the design.

By assumption, designs are patentable and therefore excludable in the sense that a firm cannot produce a durable if it does not own the design for it. The new design also contributes to the general stock of design knowledge that researchers work with when they attack new problems, and these benefits are not excludable. Designs are therefore not completely excludable.

Spillovers of design knowledge are assumed to be present because they are relevant; problems of incomplete appropriability are by all accounts very real. Nonetheless, the spillovers are not crucial. What underlies the results in the model is the assumption of nonrivalry, which implies that firms are not price takers and that price ratios are not always equal to marginal rates of transformation. Partial excludability could be extended to complete excludability without changing the basic implications of the model.[16]

Instead of relying on a single capital stock, K, output depends on an infinite list of possible types of durable intermediate inputs. If this

15. Romer (1990).

16. An example of a related equilibrium with no spillovers is given in Romer (1987a).

list is denoted by $x = \{x_i\}_{i=1}^{\infty}$, then x_{18} could represent forklifts used in the production of final output and x_{27} the number of personal computers, each measured in common units of production cost.

Final output is assumed to be a function of the list of producer inputs, human capital, H, and physical labor, L. Following Wilfred Ethier, I represent the basic production structure for employing existing and potential future inputs by the kind of additively separable function used on the preference side by Avinash Dixit and Joseph Stiglitz:[17]

$$(1) \qquad Y(L, H, x) = g(H, L) \sum_{i=1}^{\infty} x_i^{\gamma}.$$

At time t, a firm will be able to use only those durables that have already been invented. If $A(t)$ denotes the number of different types of goods that have been designed by time t, $x(i) = 0$ for any $i > A(t)$. Thus at every date, only a finite number of the terms in the sum are other than zero.

Conventional production functions aggregate together all different types of producer durables into a single aggregate capital input, $K = \Sigma_i x_i$. Then the sum K is raised to a power such as γ. This implies that all the different types of capital goods are perfect substitutes for each other; one additional dollar's worth of capital in the form of a forklift has the same effect on the marginal productivity of personal computers as an additional dollar's worth of personal computers. In the specification used here, different types of capital have additively separable effects on output. The personal computers have a marginal product that is independent of the number of forklifts in use. Even if there are many forklifts in use, the marginal product of the first personal computer is very high. As a result, it is important to distinguish the growth in total capital that comes from adding units of the existing set of durables and the growth in capital that comes from bringing new types of durables into use. The first type of capital accumulation is associated with the usual diminishing returns to capital accumulation. The second is not.

The subfunction $g(H, L)$ is assumed to be homogeneous of degree $1 - \gamma$ so that the function Y is homogeneous of degree 1. In my previous paper, g was a Cobb-Douglas function.[18] Here it can take on the general constant-elasticity-of-substitution form:

17. Ethier (1982); and Dixit and Stiglitz (1975).
18. Romer (1990).

(2) $$g(H, L) = [\alpha H^{\beta} + (1 - \alpha)L^{\beta}]^{(1-\gamma)/\beta}, \beta \epsilon (-\infty, 1].$$

In these expressions, human capital, H, should be interpreted as the total amount of human capital in the work force, not the average level. In a heterogeneous pool of workers, it would refer to the total number of person-years of education that have been obtained by workers in the pool. Physical labor would then be measured by a count of bodies. Alternatively, human capital, H, could refer to the number of skilled workers and physical labor, L, to the number of unskilled workers. Under either interpretation it is clear that L and H can vary separately and that replicating a given pool of workers requires that both L and H double.

Production of Intermediate Producer Durables

For any durable that already has a design, units of the durable can be produced at a constant unit cost in terms of forgone output. Let K be an accounting measure of forgone consumption that, except for depreciation, is constructed just as aggregate capital measures are:

(3) $$\dot{K} = Y - C.$$

In this accounting measure, no allowance is made for depreciation because the underlying durables are assumed not to depreciate. Adding exponential depreciation would introduce a familiar term to the user cost of durable goods and would otherwise leave the analysis unchanged. The aggregate, K, is simply an accounting device that measures the consumption-good value of the total stock of all of the durables $x(i)$ available.

As in the one-sector growth model, the model here assumes that final output goods can be converted one-for-one into units of durable capital goods. As in that model, this assumption is based on the idea that the inputs freed from the production of final output when consumption is reduced are put to use in the sector that produces capital goods.

In the analysis that follows, it is useful to keep track of a parameter that reflects the cost in terms of forgone consumption of one unit of any durable. Thus suppose that it takes η units of output to produce one unit of any of the different types of producer durables. Given the assumption that the durables do not depreciate and given the definition of the accounting measure, K, it follows that at every date, $K = \eta \Sigma_i x_i$.

Because durables do not suffer any depreciation, they produce a constant flow of services forever. Inserting x_i into the production function is the usual abuse of notation whereby the service flow from a stock of durables is equated with the stock itself.

Production of New Designs

A realistic treatment of the process of discovering and designing a new input in production would take account of the fact that new discoveries are indivisible goods and that the technology for producing them is subject to considerable uncertainty. The assumption made here is that these features need not be given primary emphasis in an analysis of technological change at the aggregate level over long time intervals. (They would of course be very important at the microeconomic level and in the short run.) Thus the output of new designs is assumed to be a deterministic function of the inputs used in research. To avoid integer constraints, the index, i, is treated as a continuous variable. Formally, the summation in equation 1 must be reinterpreted as an integral, but this has no substantive implications.

With these simplifying assumptions, the rate of increase in the number of designs, A, is written as

$$\dot{A} = \delta H_A A. \tag{4}$$

In this equation, H_A denotes the amount of human capital devoted to research. Human capital devoted to producing current output is henceforth denoted by H_Y. Total human capital in the population is the sum, $H = H_A + H_Y$. The specification for the manufacturing sector given above implicitly assumes that the factor intensities are the same in the production of final output and in the production of durable capital goods. In contrast, equation 4 specifies that research is human-capital intensive in a very strong sense. Physical labor and producer durables are assumed not to help at all in research. This is, of course, an exaggeration but one that is worth the simplification it permits.

The main features of this specification are that the rate of production of new designs is increasing in the amount of human capital employed and that the productivity of a unit of human capital is increasing in the total number of designs that currently exist. That the dependence on H_A is positive is the fundamental assumption in the model. If changes in research effort did not affect the rate of production of new designs, there would be no reason to exert the effort. The growth rate would be exogenously fixed, just as it is in the neoclassical model. Making the dependence on H_A linear simplifies the analysis but is not essential for the qualitative results. A more realistic formulation that recognizes some of the difficulties of coordination inherent in a more extensive research effort might make research output depend on H_A raised to a power less than one. This would not change the qualitative behavior of the model but would complicate the algebra. Because the model here is used only to sign the effects that various interventions have on the rate of growth, the exact functional form is of little importance.

The specification in equation 4 captures the cumulative nature of knowledge by making the productivity of human capital in research an increasing function of A. Because human capital is measured as it is in growth accounting or labor economics—in years of education—an engineer with 24 years of education has the same amount of human capital regardless of when he or she works. Equation 4 implies that the engineer working today is more productive than an engineer of 100 years ago because the modern engineer can make use of a much broader array of information to solve design problems.

Making the dependence on A linear is important for the existence of a balanced growth equilibrium in which all variables grow at a constant rate. Without the assumption that $\dot{A}$ is linear in A, the analysis would still be feasible, but it would be much harder. For example, if the productivity of engineers eventually stops growing as the set of all possible ideas is exhausted, the dependence on A should eventually be less than linear. In this case the rate of growth would eventually go to zero. Studying this kind of model is feasible in principle, but it would require explicit attention to dynamic paths along which growth rates vary, a kind of analysis much harder to undertake than balanced-growth analysis. What happens in the far future for values of A that are very large relative to the current values has very little effect on the current behavior of the economy. In the absence of any evidence that opportunities in science and engineering are petering out, the linearity assumption should not be too misleading for near-term analysis.

Characterization of an Equilibrium

Because any firm can bid for a new design, a researcher who produces a design can extract the present discounted value of the monopoly rent associated with the new good. The fixed design cost (that is, the nonrival design input) induces the only departure from price taking in this model. All markets other than the rental market for each specific producer durable are characterized by price-taking competition.

Because consumers care only about the single final output good, their preferences are needed only to determine interest rates. The specification used here is familiar,

$$U[C(\cdot)] = \int_0^\infty e^{-\rho t} \frac{C(t)^{1-\sigma}}{1-\sigma}\, dt.$$

These preferences are used only to derive the intertemporal optimization condition that relates interest rates to preferences and the rate of growth of consumption:

$$r = \sigma \frac{\dot{C}}{C} + \rho. \tag{5}$$

What really matters in the analysis that follows is the technology and the imperfectly competitive market structure that results from the nonrivalry of designs. Any other rule relating interest rates to consumer choices could be substituted for equation 5 without changing the thrust of the results.

The intuition behind the equilibrium behavior of this model can be inferred from equations 1 and 3. From the symmetry between all of the different types of producer durables, it follows that all of them in existence at time t will be used at the same level $\bar{x}$. From equation 3, it follows that $\bar{x}$ is related to A and K by $\bar{x} = K/(\eta A)$. Substituting this expression into equation 1 and using the fact that the range of available durables at time t is $A(t)$, it follows that output can be written as

$$Y = g(H, L) \int_0^A x(i)^\gamma di = g(H_Y, L)A \left(\frac{K}{\eta A}\right)^\gamma .$$

Using the fact that $g(\cdot)$ is homogeneous of degree $1 - \gamma$, this can be written as

$$Y = g(AH_Y, AL)K_y^{\gamma-\gamma}.$$

Formally, A behaves just like labor- and human-capital-augmenting technological change. Thus in balanced growth, the rate of growth of A is equal to the rate of growth of output. By equation 4, it follows that the amount of human capital, H_A, devoted to research determines the rate of growth of A.

At any time, the allocation of human capital between manufacturing and research is determined by the requirement that the return to human capital be the same in both the manufacturing and research sectors. In manufacturing, which is competitive, the wage paid for human capital is $w_H = \partial Y / \partial H$:

$$w_H = D_1 g(H, L) A \left(\frac{K}{\eta A}\right)^{\gamma}. \tag{6}$$

In the research sector the wage for human capital can be expressed as $w_H = P_A \delta A$. Because each researcher is free to exploit all the knowledge, A, all of the revenue from the sale of designs is captured by researchers.

It remains to determine the value of a patent. This is just the present discounted value of the monopoly profit that the holder of the patent can extract. If H_Y is the total human capital used in producing final output and L is total labor, the aggregate derived (inverse) demand curve for durable good i can be expressed as

$$p(x) = g(H_Y, L)\gamma x^{\gamma - 1}. \tag{7}$$

Using consumption goods as numeraire, the cost of building x units of durable good i is ηx. At any time t the rental cost of the raw capital is $r\eta x$. The present value of monopoly rent that can be extracted is then

$$P_A = \max_{x(\cdot)} \int e^{-\int_0^t r(s)ds} \{p[x(t)]x(t) - r(t)\eta x(t)\}\, dt. \tag{8}$$

Formally, this statement of the problem of the monopolist assumes that $x(\cdot)$ can be varied over time and that when the amount of x supplied falls, the raw capital used to construct the previous level of x can be released. That is, the capital used to produce the individual durables is putty-putty. (The research costs are of course sunk costs.) In fact, because of all of the stationarity built into this model, the optimal monopoly level of $\bar{x}$ and the optimal price $\bar{p} = p(\bar{x})$ do not vary with time in equilibrium. As a result, nothing would change if capital were instead putty-clay.

What makes $\bar{x}$ constant is that the interest rate, r, the level of human capital in manufacturing, H_Y, and labor, L, are constant. The total level of L is fixed by assumption, and L has no alternative uses in this specification. (The effects of different levels of L will be inferred from an exercise in comparative dynamics in which two balanced-growth paths with different constant levels of L are compared.) The total amount of human capital, H, is also assumed to be constant. The constancy of both H and L is assumed simply to facilitate the balanced growth analysis.

It remains to show that a constant value of H_Y and r will be consistent with the requirements for equilibrium along a balanced growth path. If H_Y and r are constant, the expression for the present discounted value of the monopoly rent is

$$P_A = \bar{x}\eta \frac{1-\gamma}{\gamma}, \tag{9}$$

with $\bar{x}$ determined by

$$\bar{x} = \left(\frac{g(H_Y, L)\gamma^2}{r\eta}\right)^{\frac{1}{1-\gamma}}. \tag{10}$$

These results follow easily because the inverse demand curve is isoelastic.

The expressions for the wage, w_H, implied by opportunities in the research sector can be combined with the expression from the manufacturing sector in equation 6 as follows. The return earned by human capital devoted to research is the number of designs produced per unit time per unit human capital, δA, times the price, P_A. If the right side of equation 6 is equated to $\delta P_A A$ and the expression substituted for $\bar{x}$ in terms of P_A from equation 9, and if terms in P_A are collected on the left, then

$$P_A^{1-\gamma} = \frac{1}{\delta} D_1 g(H_Y, L) \left(\frac{\gamma}{1-\gamma}\right)^{\gamma} \eta - \gamma. \tag{11}$$

Then, if the expression for $\bar{x}$ from equation 10 is substituted into equation 9, and both sides are raised to the power $1 - \gamma$, one gets a second equation in $P_A^{1-\gamma}$:

$$P_A^{1-\gamma} = \frac{1}{r}\gamma^2 g(H_Y, L)\left(\frac{1-\gamma}{\gamma}\right)^{1-\gamma}\eta - \gamma. \tag{12}$$

Equating the right sides of equations 11 and 12 and solving for r gives

$$r = \delta\gamma(1-\gamma)\frac{D_1 g(H_Y, L)}{g(H_Y, L)}. \tag{13}$$

This expression is easier to interpret first in the case in which the elasticity of substitution parameter, β, in the functional form for g takes on the value 0, so that g reduces to the Cobb-Douglas form

$$g(H_Y, L) = H_Y^{\alpha(1-\gamma)}L^{(1-\alpha)(1-\gamma)}. \tag{14}$$

Then equation 13 reduces to

$$r = (\delta\gamma/\alpha)H_Y.$$

The theory can now proceed in one of two ways. Either preferences can be treated as being such that equation 5 holds, so that $r = \sigma(\dot{C}/C) + \rho$, or preferences can be left unspecified, and r can simply be taken as a parameter of the model that is unaffected by changes in the rate of growth of output. In the first case, the solution for H_A follows from the fact that in balanced growth $\dot{C}/C$ must be equal to $\dot{A}/A = \delta H_A = \dot{Y}/Y$. In this case these two linear expressions in H_Y and r can be solved to yield a resulting expression for human capital in research,

$$H_A = \frac{H - (\alpha\rho/\gamma\delta)}{1 + (\sigma\alpha/\gamma)}. \tag{15}$$

Since the growth rate of output and consumption is δ times H_A, this expression is subject to parameter restrictions necessary to ensure that the integral of discounted future utility is finite. Specifically, $(1-\sigma)[\delta H - (\alpha\rho/\gamma)]/[1 + (\alpha\sigma/\alpha)]$ must be less than the discount rate ρ. This is automatically true if σ is greater than 1. If the interest rate is taken as a parameter, the comparable expression for human capital in research is

$$H_A = H - \frac{r\alpha}{\delta\gamma}. \tag{16}$$

The basic features of interest in equations 15 and 16 are the same. First, human capital in research increases with total human capital, H.

This has two implications. In a closed economy an increase in total human capital will lead to a more than proportional increase in the amount of human capital devoted to research. Consequently, combining two isolated economies into a single integrated economy will lead to an increase in the worldwide rate of amount of human capital in research and therefore to an increase in the worldwide rate of growth. There are some subleties about the general effects that increased trade has on the allocation of human capital to research. Some of these issues are considered by Luis Rivera-Batiz and me for the case of trade between symmetric countries.[19] The more complicated effects that can arise when trade is introduced between countries that differ in some crucial way are explored by Gene Grossman and Elhanan Helpman.[20] But the basic idea here is clear. The value of incurring a fixed cost depends on the size of the market in which the resulting good can be sold, and it is efficient to ensure that human capital in different countries is not engaged in redundant design efforts. With free trade it is possible to put to better use human capital that might otherwise be used to redesign the wheel.

The second feature of these equations is that a decrease in the interest rate is associated with an increase in human capital devoted to research and an increase in the growth rate. In equation 16 this is immediately clear. In equation 15 the decrease follows because a fall in either of the preference parameters, ρ or σ, will reduce the interest rate and increase the growth rate. Because the equations here refer to balanced-growth paths, the calculated effects are long-run effects. In the transition to a new balanced-growth-rate path, output might appear to fall if the output of the human capital shifted into research is not counted as part of total GDP, as often happens in practice.

In contrast with the clear growth effects of changes in interest rates or preference parameters, neither total labor, L, nor the magnitude of the parameter η affects H_A in equations 15 and 16. Consequently, neither has any effect on the rate of growth. As noted earlier, the parameter η is of interest because it represents the cost in terms of consumption goods of the units of the producer durables that are produced. A direct subsidy to capital accumulation—for example, an investment tax credit

19. Rivera-Batiz and Romer (1989).
20. Grossman and Helpman (1989a, 1989b).

financed from lump-sum taxes—will have an effect on the equilibrium rate of growth that is the same as a change in η. Therefore an investment tax credit has no long-term effect on the rate of growth.

The intuition for these invariance results depends on recognizing that the research sector must always compete with the manufacturing sector for human capital. A subsidy for capital accumulation that decreases η affects returns to human capital in both sectors. From equation 10 it follows that a reduction in η leads to an increase in $\bar{x}$. This increases the marginal product of human capital in manufacturing, but it also increases the demand for durables and therefore the returns to human capital in research. For the functional form used here, these effects exactly cancel one another. Consequently, a subsidy for physical capital accumulation has very different effects on the rate of growth than an intervention that lowers the interest rate.

An increase in L has similar effects. First, it directly increases the marginal product of human capital used in manufacturing. Second, it increases the marginal product of each of the producer durables and thereby increases the monopoly rent that a researcher who discovers a new design can extract. This causes the price for a design to go up, which raises the return to human capital in research. With the Cobb-Douglas form for the function g imposed in equation 14, these two effects exactly cancel. The increase in L has no effect on the allocation of H.

Because this result depends on the exact cancellation of two offsetting effects, it will not be robust to small changes in the specification. This can be shown by letting β from equation 2 take on values different from 0. In this case, equation 13 becomes

$$(17) \quad r = \delta\gamma H_Y + \frac{\delta\gamma(1 - \alpha)}{\alpha} L^{\beta} H_Y^{1-\beta}$$
$$= \delta\gamma(H - H_A) + \frac{\delta\gamma(1 - \alpha)}{\alpha} L^{\beta}(H - H_A)^{1-\beta}.$$

Whether one substitutes the expression $(\sigma\dot{C}/C) + \rho = \sigma\delta H_A + \rho$ for the interest rate r or merely treats r as a parameter, this equation cannot be explicitly solved for H_A. It is nonetheless easy to verify that if H and L are complements (that is, if β is less than 0), an increase in L leads to a decrease in H_A and in the long-run rate of growth.

H_A falls when L goes up because the elasticity of substitution between labor and human capital is smaller than the elasticity of substitution between physical capital and the composite of L and H. For example, this will hold if workers, L, and managers, H, are used in fixed proportions and if it is possible to use fewer workers and managers by installing more capital.

In this environment an increase in L has a larger positive effect on the marginal productivity of managers than it has on the marginal product of producer durables. An increase in L therefore causes the returns to human capital in management to increase by more than the returns to human capital in research, and human capital will shift from research into management. The negative relation between L and H_A would presumably still obtain if the elasticity of substitution between labor and human capital were left the same as in the Cobb-Douglas case and the elasticity of substitution between labor and physical capital were reduced as well. For the functional forms used here, the first modification is much easier to implement. Finally, because η does not appear in equation 17, a subsidy for capital still has no effect on the amount of human capital allocated to research or on the long-run rate of growth.

Summary of Theoretical Results

The two basic ideas in the model are that the fixed cost of doing research must be covered by a stream of revenue that arises in the future and that human capital, which is the primary input in research, has alternative uses in the direct production of output. From these ideas it follows that an increase in the stock of labor can cause human capital to shift from research into production of final goods if in the production of final goods labor is a better substitute for physical capital than it is for human capital. In the long run this leads to a fall in the rate of growth. It also follows that the marginal product of capital is conceptually distinct from the market rate of interest. It is the interest rate that influences the decision to incur a fixed cost. A subsidy for physical capital accumulation can have offsetting effects on the incentive to do research. As a result, a subsidy could increase the share of GDP devoted to investment without changing the market rate of interest, the total research effort, or the long-term rate of growth.

Evidence

The fit between the variables in the model and the ones for which data exist is not tight. The data used here contain no measure of subsidies for capital, no proxy for η. For the countries in this sample, only a measure of the investment share is available. This is not a serious problem because variation in η should induce variation in the investment share. The theoretical analysis also compares two different constant levels of labor, L, whereas the data refer to different rates of growth of labor. Again, this is not a problem for the kind of informal data analysis undertaken here. The same mechanism that causes an increase in L to drive human capital out of research will operate if L or, more realistically, both L and H grow. The faster the rate of growth of L, the bigger this effect should be. Modeling this explicitly does not appear to be feasible within the narrow confines of balanced-growth analysis but would be feasible in a complete dynamic analysis.

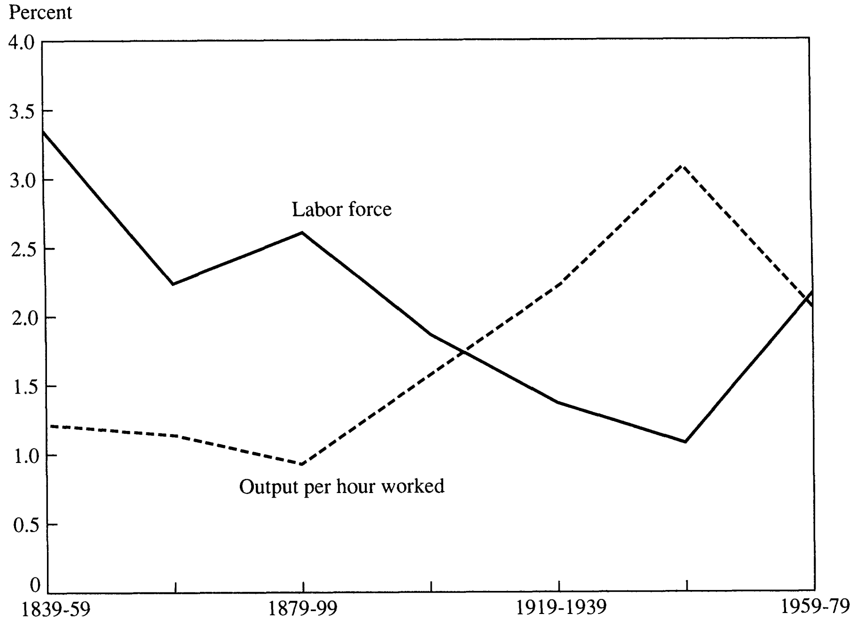
With these caveats in mind, it should be clear that there is no reason to attempt formal structural estimation or hypothesis testing with the data presented here. What follows is an attempt to summarize features of the data that are suggested by the theory.

Labor

Figure 1 presents data on the long-run relationship between growth in the labor force and growth in output per hour worked for the United States.[21] Average annual growth rates for these two variables are taken over periods of 20 years to remove as much of the variation in the business cycle as possible. At business cycle frequencies, it is well known that labor productivity goes down when output and hence employment go down. In the figure the opposite relationship is evident at

21. The data used here stay as close as possible to the data on the private business sector in the postwar period that are available from the Department of Commerce. From 1890 to 1950, data for the private business sector are taken from Kendrick (1961). Before 1890, data on employment are from Lebergott (1966). For the 1870s and 1880s data on average hours worked are from Kendrick (1961). For the 1840s, 1850s, and 1860s hours worked were assumed to remain constant. Before 1890, data on output are from Gallman (1966) augmented by unpublished data from Gallman's original worksheets (used with permission). Except for the use of the data from the worksheets, which influence only the value for output per hour worked in the period from 1879 to 1899, this figure is the same as the one that appeared in Romer (1987b).

Figure 1. Labor Force and Productivity Growth, Selected Periods, 1839–1979



Sources: See note 21.

low frequencies. It persists throughout the century and a half for which data exist. The recent negative correlation between growth in the labor force and growth of labor productivity is not a new phenomenon.

Ideally, what one would like to observe is an exogenous change in the size or rate of growth of the labor force and then trace through its effects on output and productivity. In this figure, growth in labor productivity is compared with growth in the labor force rather than with growth in hours worked, because the labor force variable should be closer to being exogenous. (The main difference between growth in hours worked and growth in the labor force occurs during the period that includes the 1930s, when hours worked grew much less rapidly than the labor force.) At these very low frequencies, movements in the labor force are dominated in the early period by changes in immigration and in the later period by the baby boom and the increase in the labor force participation rate for women.

This figure shows that trend growth in output per hour worked is negatively related to growth in labor supply, not positively related as it is at business cycle frequencies. Moreover, the magnitude of the variation is much larger than one would expect based on neoclassical assumptions. In the long run, one would expect the rate of growth of the capital stock to increase by the rate of growth of the labor force. If it did, output would increase one-for-one with the labor force, and labor productivity growth would not be affected by the rate of growth of the labor force. A 20-year period may not be long enough to capture the full long-term effect, but even in the very short term, in which the rate of growth of capital does not adjust at all in response to movements in labor supply, a 1 percent increase in the rate of growth of labor supplied should lead to at most a 0.3 percent reduction in the rate of growth of output per hour worked. In the figure the reduction is closer to one-for-one than to three-for-one. The explanation suggested by the model is that research effort responds to incentives. During periods when labor is growing rapidly relative to the stock of human capital, the model predicts that the returns to human capital increase relative to those of labor and increase by more in the production of final goods. As a result, a smaller fraction of total human capital is devoted to research.

Superficially, this description bears some resemblance to developments in the United States during the past 20 to 30 years. The kind of human capital that is relevant for the research activities described here is postgraduate education in applied science and engineering (not in basic scientific research that takes place in universities), and there is widespread concern that the quality and quantity of the new entrants to these professions have been decreasing. Citizens of the United States and other native English speakers perceive that returns to education are higher in other fields, such as law, medicine, and management. Graduate programs are increasingly populated and staffed by foreign nationals who apparently have difficulty gaining access to employment in the higher-paying occupations chosen by citizens of the United States. Especially in the past decade, this change has occurred when the wage of skilled human capital relative to labor has been increasing, just as the model would predict.

Nor is it implausible that these effects could operate over a 20- or 30-year horizon. Evidence from industries such as machine tools and consumer electronics suggests that during this time span, the United States has fallen from a position of clear superiority in engineering to one of serious weakness. This change is most dramatic in the automotive industry. Even a leading firm like Ford, which has made substantial progress in manufacturing processes, has given up any attempt to remain competitive in designing small cars. Current plans call for Ford to manufacture and sell its next small car, but the design has already been contracted out to Mazda.

There is reason to expect that in the nineteenth century, the supply response to incentives to innovate was even larger than it is now. Kenneth Sokoloff shows that patenting activity in the first half of the century was responsive to access to large markets.[22] In subsequent work Sokoloff and Zorina Khan trace the patenting activity of individual inventors and show that the patenting activity of an important group of individuals is very responsive to short-term economic incentives.[23] In the years covered by this work, valuable inventions could be created by people with general mechanical knowledge, and one would expect that the response of an increase in the returns to research would be more rapid than it would be now. Entry and exit, however, are not symmetric. There is no reason to believe that exit from engineering is any slower now than it was in the past.

The time-series finding in the United States is consistent with the differential between productivity growth in the United States and Europe. My earlier paper documented the sharply divergent behavior of growth in hours worked in the United States and Europe that was matched by very different productivity growth rates.[24] The same finding has been emphasized by Richard Freeman: the flip side of the very low rate of growth of employment and hours worked in Europe compared with that in the United States is a much higher rate of growth of productivity.[25] One obvious testable implication of this model is that the relative wage of people engaged in engineering in Europe is higher than it is in the United States.

22. Sokoloff (1988).
23. Sokoloff and Khan (1990).
24. Romer (1987b).
25. Freeman (1989).

Growth and Investment

The theoretical exercise in the specification of the technology showed that even if rapid growth in the labor force is bad for research and technology growth, this does not mean that rapid growth of capital is good for research and growth. Tables 1 through 4 describe the evidence across countries on the correlation between the rate of growth of per capita income and the ratio of total investment (in both the public and private sector) to GDP. As noted earlier, this is exploratory data analysis, not structural estimation or formal hypothesis testing.

The model shows that the correlation between the rate of growth and the investment share depends crucially on the source of the variation. If a higher investment share is induced by a subsidy or some other policy that leaves invariant the rate of growth of the technology, then the rate of growth of output will not be affected in the long run. The marginal product of capital will fall because of the usual diminishing

Table 1. Relationship of Growth Rate of Per Capita GDP and the Investment Share (Dependent Variable: *GROWTH*)

Independent variable[a]	*Coefficient*	*Standard error*	*t-statistic*	*Two-tail significance*
C	2.20	0.79	2.78	0.007
POP GROW	0.968	0.21	4.71	0.000
Y AVG	−0.00025	9.6E−0.5	−2.55	0.013
INV	0.182	0.027	6.67	0.000
GOV	−0.099	0.028	−3.57	0.001
AF DUMMY	−1.27	0.41	−3.13	0.002
LA DUMMY	−1.24	0.42	−2.96	0.004

Summary statistics
$\bar{R}^2$ 0.449
Adjusted $\bar{R}^2$ 0.417
Standard error of regression 1.436
Mean of dependent variable 4.07
Standard deviation of dependent variable 1.88
Number of observations 112

Source: See note a; and unpublished data from the World Bank.

a. *GROWTH*: The rate of growth of per capita GDP from *RGDP*2 in Summers and Heston (1988). In percent per year times 100. Mean 1.96; range 6.65 to −2.88. *POP GROW*: the average annual rate of growth of the population measured from 1960 to 1985. In percent times 100. Mean 2.1; range 4.3 to 0.3 (Summers and Heston). *Y AVG*: The geometric average of real per capita income in 1960 and 1985, in 1980 dollars. Raw data from *RGDP*2 in Summers and Heston. Mean 2389; range 9630 to 255. *INV*: the ratio of current-price investment (public and private) to current-price GDP times 100. From *CI* in Summers and Heston. Mean 14.8; range 29.1 to 3.0. *GOV*: the ratio of current-price noninvestment spending by the government to current-price GDP. From *CG* in Summers and Heston. Mean 16.1; range 31.1 to 4.0. *AF DUMMY*: dummy variable for Africa. *LA DUMMY*: dummy variable for Latin America, including South America, Central America, and Mexico.

Table 2. Relationship of Growth Rate of Per Capita GDP, the Investment Share, and the Square of the Investment Share (Dependent Variable: *GROWTH*)[a]

Independent variable[b]	*Coefficient*	*Standard error*	*t-statistic*	*Two-tail significance*
C	0.949	0.947	1.00	0.318
POP GROW	0.885	0.204	4.33	0.000
Y AVG	−0.00027	9.46E–05	−2.83	0.006
INV	0.422	0.107	3.93	0.000
INV SQ	−0.0076	0.0032	−2.31	0.023
GOV	−0.110	0.028	−3.99	0.000
AF DUMMY	−1.04	0.41	−2.53	0.013
LA DUMMY	−1.25	0.41	−3.04	0.003

Summary statistics
$\bar{R}^2$ 0.476
Adjusted $\bar{R}^2$ 0.440
Standard error of regression 1.41
Mean of dependent variable 4.07
Standard deviation of dependent variable 1.88
Number of observations 112

Sources: See Table 1, note a.
a. Sample statistics for the estimated coefficient of *I/Y*: *mean*: 0.309; *standard deviation*: 0.049; *maximum*: 0.399; *minimum*: 0.201. For each country the estimated coefficient of *I/Y* is the sum of the coefficient estimated for *INV* plus the level of *INV* times the estimated coefficient of *INV SQ*.
b. *INV SQ*: *INV* squared. For other definitions and scale information needed to interpret coefficients, see table 1, note a.

returns. But if the underlying source of variation is variation in the rate of growth of the technology induced by differences in the stock of human capital or the degree of integration with world markets, then increased growth in the technology will lead to faster growth in both output and capital. Because the rate of growth of capital will be higher and the capital output ratio (determined by $\bar{x}$) will be the same, the investment share will have to be higher.

Recall that production for the model can be written in terms of A and K as

$$Y = g(H_Y, L)A\left(\frac{K}{\eta A}\right)^{\gamma} = g(H_Y/L, 1)L^{1-\gamma}A\left(\frac{K}{\eta A}\right)^{\gamma},$$

where the second equality follows because the function $g(\cdot)$ is homogeneous of degree $1 - \gamma$. If one takes logarithms of both sides and then takes time derivatives, and lets variables with a circumflex denote rates of change, this gives

Table 3. Relationship of Investment Share of GNP and Level of Income and Exports (Dependent Variable: *INV*)

Independent variable[a]	*Coefficient*	*Standard error*	*t-statistic*	*Two-tail significance*
C	8.00	1.07	7.48	0.000
Y AVG	0.0014	0.00018	7.52	0.000
EXP	11.00	3.13	3.51	0.001

Summary statistics
$\overline{R}^2$ 0.473
Adjusted $\overline{R}^2$ 0.461
Standard error of regression 4.71
Mean of dependent variable 15.0
Standard deviation of dependent variable 6.42
Number of observations 90

Sources: See Table 1, note a.
a. *EXP*: the average of the ratio of exports to GDP for 1960–85 (unpublished data from the World Bank). Mean 27.7; range 85.5 to 5.1. For other definitions and scale information needed to interpret the coefficients, see table 1, note a.

$$(18)\quad \hat{Y} = \frac{d}{dt}\log[g(H_Y/L,\ 1)] + (1-\gamma)\hat{L} + (1-\gamma)\hat{A} + \gamma\hat{K}.$$

If Δ denotes the exponential rate of depreciation, then the ratio of gross investment, I, to total output, Y, is related to $\hat{K}$ by $\hat{K} = [(I/Y)(Y/K)] - \Delta$. Substituting this into equation 18 gives

$$(19)\quad \hat{Y} = \frac{d}{dt}\log[g(H_Y/L,\ 1)] + (1-\gamma)\hat{L} + (1-\gamma)\hat{A} + \gamma\frac{Y}{K}\frac{I}{Y} + \gamma\Delta.$$

The model from the last section predicts that variation in I/Y does not affect $\hat{A}$ but that $\hat{L}$ has a negative effect on $\hat{A}$. Using these predictions and equation 19, one may make a back-of-the-envelope calculation of what a regression of $\hat{Y}$ on I/Y and a proxy for $\hat{L}$ should yield. Assuming that H_Y/L does not change too much during the sample period, the crucial omitted variable is $\hat{A}$. If one had observations on $\hat{L}$ itself, the coefficient on $\hat{L}$ would include the sum of the true coefficient, $1-\gamma$, and the projection of $\hat{A}$ on $\hat{L}$ times $1-\gamma$. For this production function, $1-\gamma$ should be approximately equal to one minus the share of capital in total income, roughly 0.7. Under the arguments suggested above, the projection of $\hat{A}$ on $\hat{L}$ should be negative, so the estimated coefficient

Table 4. Relationship of Growth, the Investment Share, and Interaction Terms with the Investment Share (Dependent Variable: *GROWTH*)[a]

Independent variable[b]	*Coefficient*	*Standard error*	*t-statistic*	*Two-tail significance*
C	0.99	0.87	1.14	0.258
POP GROW	1.21	0.22	5.42	0.000
Y AVG	0.00055	0.00027	2.00	0.049
INV	0.224	0.044	5.13	0.000
INV × *Y AVG*	−3.64E−05	1.36E−05	−2.67	0.009
INV × *EXP*	−0.00018	0.00049	−0.37	0.711
GOV	−0.108	0.030	−3.65	0.001
AF DUMMY	−1.42	0.45	−3.16	0.002
LA DUMMY	−1.54	0.43	−3.58	0.001

Summary statistics
$\bar{R}^2$ 0.537
Adjusted $\bar{R}^2$ 0.491
Standard error of regression 1.28
Mean of dependent variable 3.87
Standard deviation of dependent variable 1.80
Number of observations 90

Sources: See Table 1, note a, and Table 3, note a.

a. The same regression as in table 3, except that the square of *INV* has been replaced by interaction terms between *INV* and the two variables that have explanatory power for *INV* in table 3.

Sample statistics for the estimated coefficient on *I/Y*: *mean*: 0.124; *standard deviation*: 0.093; *maximum*: 0.211; *minimum*: −0.129. For each country the estimated coefficient is the coefficient estimated for *INV*, plus the level of *Y AVG* times the estimated coefficient on *INV* × *Y AVG*, plus *EXP* times the estimated coefficient on *INV* × *EXP*.

b. For definitions and scale information needed to interpret the coefficients, see table 1, note a, and table 3, note a.

would be less than this. Moreover, since population growth is an imperfect proxy for growth in the labor force, the coefficient may also be biased toward 0 by the usual measurement-error argument.

The interpretation of the coefficient on I/Y depends on what assumptions one makes about the covariance between $\hat{A}$ and $\hat{K}$. Suppose that all the exogenous variation comes from A and that K responds passively; that is, there is no variation across countries in such policy interventions as subsidies to capital accumulation. Then in the long run $\hat{A}$ and $\hat{K}$ should be very closely correlated, and there is no force that induces differences in the output to capital ratio Y/K (or equivalently, the marginal product of capital) across countries. A regression of $\hat{Y}$ on $\hat{K}$ would find a coefficient on $\hat{K}$ of approximately 1. Therefore, a regression of $\hat{Y}$ on I/Y should have a coefficient of roughly Y/K, or about 1/3. If both $\hat{L}$ and I/Y are included, the coefficient on $\hat{L}$ should be biased down from $1 - \gamma$ and the coefficient on I/K should be biased up from $\gamma Y/K$. If $\hat{A}$ and I/Y are more closely correlated than $\hat{A}$ and $\hat{L}$, for example because of measurement error in $\hat{L}$, then most of the bias will fall on the coefficient of I/Y.

Alternatively, if it is I/Y that varies independently and the model is correct in its prediction that subsidies to capital accumulation do not by themselves increase the rate of growth of the technology, then in the long run the variation in I/Y across countries should be exactly offset by variation in the output-capital ratio, Y/K, and the marginal product of capital, which is proportional to Y/K. This follows because $\hat{K}$ is equal to $(I/Y)(Y/K - \Delta$ by definition, and $\hat{K}$ must be equal to $\hat{A}$ in the long run. (This of course is Solow's original insight about the effects of variation of I/Y.) In this case the estimated coefficient on I/K should be zero.

Both these assumptions are modified somewhat if the 25-year interval used here is too short for the full adjustment to the long run to have taken place. In the first case, suppose that just before the beginning of the observation period $\hat{A}$ increases. If 25 years is too short a period for I/Y to respond fully to the increase in the rate of growth of the technology, the projection of $\hat{A}$ on I/Y will be larger than in the case of full adjustment and the value of Y/K will be larger for this country than for others. In the limit, as the change in I/Y in response to the increase in $\hat{A}$ becomes smaller, this projection increases without bound. For both reasons, the estimated coefficient on I/K will be even larger than it would be in the long-run case. In the case in which I/K varies independently, in the very short run the output-to-capital ratio does not have a chance to adjust, and the increase in output should be $\gamma(Y/K)$; that is, the short-run coefficient should be equal to the marginal product of capital. If Y/K is 1/3 and γ is 0.3, then the estimated coefficient should be approximately 0.1.

In summary, when the exogenous variation is in $\hat{A}$, the estimated coefficient on I/Y should lie in the range $(Y/K, \infty) \cong (1/3, \infty)$. If the variation in I/Y is entirely independent of $\hat{A}$, the estimated coefficient should lie in the range $(0, \gamma Y/K) \cong (0, 0.1)$. In each case, the longer the observation period and the closer the adjustment to the full long-run adjustment, the closer the estimated coefficient should be to the lower bound of the relevant interval.

Table 1 presents results that are typical of the large number of cross-sectional regressions that can be estimated. Variables such as the share of noninvestment government spending in GDP, the initial level of income, and continent dummy variables that have previously been found to be correlated with the rate of growth have been included here. As noted earlier, population growth is used as a proxy for labor force growth. If the population growth rate was more or less constant in the entire postwar period, this should be a reasonable approximation for the 25-year period studied here.

Many other right-hand-side variables could be considered in a regression of this kind. Some of the estimated coefficients reported here are not robust to changes in the specification, but in all specifications the investment share has an estimated coefficient that is on the order of 0.1 to 0.2 and has a t-statistic that is large, on the order of 4 to 7. Based on the assumption that the 25-year interval should be long enough for the adjustment to a steady state to be largely complete, estimates of this magnitude suggest that both of the suggested effects are present. There is independent variation in both $\hat{A}$ and in I/Y. The estimated effect is a mixture of the two calculated effects.

Given the findings cited above, the estimated coefficient on population growth is surprisingly large. The point estimate is larger than the kind of value one would predict from neoclassical theory, although it is within two standard deviations of this value. There are two possible explanations for this finding. One is that the mechanism described above, whereby an increase in L reduces the rate of technological progress, does not operate in the less developed countries that form the overwhelming bulk of observations in this sample. It is also possible that there is positive feedback from the rate of growth of income to population growth. Many of these countries have not yet passed through the demographic transition. In a comparison of these countries before and after World War II, population growth has increased with income, the opposite of the pattern observed in developed countries. Even if fertility falls so much that the expected number of surviving children stays constant, population growth will still increase for many years when there are the kind of rapid reductions in mortality that have been observed since 1945. This growth occurs as the upper part of the age distribution fills in during the convergence to the new steady-state distribution. If there is positive feedback from the rate of growth of income to population growth for either of these reasons, the estimated coefficient in this regression will be biased upward. In this case, faster growth in L could still cause reductions in $\hat{A}$, but it is masked here by the positive bias.

The significance of the continent dummy variables is a sign that other variables are important but are not being measured. The interpretation of the coefficients on government spending and especially on the average level of income per capita is difficult. It is tempting to conclude that a large share of noninvestment government spending causes a slower rate of growth, but given that government spending is endogenously determined, this is not the only interpretation of the coefficient reported here.[26]

The income measure used here is a geometric average of the initial and terminal levels of income that are used to calculate the growth rate. This would tend to induce a spurious positive correlation between the level of income and the rate of growth because countries that start out at the same level but grow faster will have a higher average level. This bias is offset by bias introduced by measurement error. If the initial level of income is reported as being too low because of transient measurement error, then the rate of growth calculated from this measurement will be too high. This measurement-error bias causes the initial level of income to be negatively correlated with the rate of growth. In separate work, I reported evidence suggesting that the negative measurement-error bias is serious if one uses the initial level of income.[27] The average used here is an attempt at a compromise, but one should not interpret the coefficient reported here with any confidence.

Within the modest goals of this informal kind of analysis, the conjecture of interest is that variation in the investment share that is not induced by increases in the rate of growth of A should have a smaller effect on the growth rate than variation that is induced by growth in A. If it were possible to split the sample into countries in which the variation in I/Y is induced by variation in A and those in which it is not, one could test for this effect by comparing the coefficient on I/Y estimated in the two samples. There is no direct evidence on which to make this kind of split, but there is other information that can be exploited. Suppose for example that countries having values of I/Y that are very high are likely to be ones in which I/Y is much larger than $\hat{A}$. For these countries, the coefficient on I/Y should be smaller than for the others. Suppose symmetrically, that countries in which I/Y is very

26. See Barro (1989, 1990) for a more detailed examination of the theoretical and empirical issues raised by government spending.

27. Romer (1989).

low are likely to be ones in which I/Y is much less than $\hat{A}$. For them, the coefficient on I/Y should be larger. To test for this effect, one could split the sample based on the value of I/Y, or, equivalently, allow for a term $(I/Y)^2$ in the regression.

Table 2 shows the results of this latter test. The coefficient on the square of the investment share is negative and is statistically significant. Note a shows that the size of the coefficient is also significant from an economic point of view. The effective coefficient on the investment share varies with the investment share from 0.4 for countries with the lowest investment share to 0.2 for countries with the highest.

Table 3 illustrates a naive attempt to break down the variation in the investment share in terms of other variables. The only variables that have explanatory power for the investment share are the average level of income and the fraction of GDP devoted to exports. Richer countries and countries that have a higher fraction of output devoted to exports invest a higher fraction of GDP in capital accumulation. Table 4 then reports the effects of removing the interaction term of the investment share with itself (that is, the square of the investment share) and replacing it with interaction terms between the investment share and the level of income, and between the investment share and the share of exports. Of the two, only the interaction with the level of income is significant, and its effect is negative, as was the effect of the squared term in the previous table. In this specification, the estimated effective coefficients on the investment share, or the estimated rates of return to capital, vary between 0.2 and -0.1. The overall level of the coefficient on I/Y is not stable across these two specifications, perhaps because the inclusion of the trade variables means that 22 countries must be dropped from the sample because of missing data. The evidence once again suggests that there is substantial variation across countries in the marginal effect of an increase in the investment share.

Taking the results from this regression at face value, increases in I/Y that are associated with a higher level of income lead to a smaller marginal product of capital and a smaller marginal effect of I/Y on the rate of growth, whereas increases in I/Y associated with increases in the share of GDP devoted to exports are not. This pattern suggests that countries that export a higher share of GDP invest more because they have a higher rate of technological change. It is clear that it is openness, not exports per se, that drives this result. In these data, exports track imports closely, and replacing exports by imports does not change the results in either table 3 or table 4. This correlation between openness is consistent with causality running from openness to technological change, as suggested by the result from the model that market size can have a positive effect on research and growth, and by the findings of Kenneth Sokoloff.

Conclusions

The main reason for undertaking general equilibrium analysis of the kind attempted here is that it offers a framework that can be used to tie many different pieces of evidence together. If productivity behavior is considered separately in the United States during the nineteenth century, in the United States since World War II, in Europe in the same period, and in developing countries for the past 25 years, it is easy to generate many hypotheses that can explain its behavior in each setting. When evidence from all these sources is taken together, generating a unified explanation becomes a more interesting and more important scientific challenge.

The overall interpretation is as follows. Applied research effort, interpreted in a broad sense, responds positively to the returns in the research sector and negatively to opportunities in other sectors. An increase in the size of the market or in the trading area in which a country operates increases the incentives for research and thereby increases the share of investment and the rate of growth of output, with no fall in the rate of return on capital. For reasons that are not clear (perhaps because of institutional or policy factors) a higher level of income seems to be associated with a higher rate of savings and investment. Because this investment is not induced by a faster rate of technological change, the higher rate of investment is associated with a lower rate of return to capital. Combined with the evidence from the United States during the 1970s, when productivity growth rates were low and investment was relatively strong, this result suggests that higher exogenous savings have little relationship with higher technological change and productivity growth. This finding is consistent with the prediction of the model that increased capital will have effects on the allocation of human capital between research and production that are largely offsetting. Put crudely, if the current problem in the United States is too many lawyers and MBAs and too few engineers, increasing the investment tax credit may not be of much help.

A more informal but diverse body of evidence suggests that labor scarcity may be good for technological change and productivity growth, at least in developed countries. (It is possible that this effect is present in less developed countries as well but is masked in the data by positive feedback between income growth and population growth.) When labor grows, the growth rate of total factor productivity should be slower and wages for labor should fall relative to wages for human capital. The crucial qualification here is that the negative correlation between labor growth and productivity growth does not hold at business cycle frequencies; something else (for example, labor hoarding or mismeasurement of labor and capital actually used) must explain why output and productivity move together over the business cycle.

The only policy conclusions that one should draw from an exercise of this kind is to do more research. What the results presented here suggest is that for this research to be productive, it must move outside the narrow confines of neoclassical growth theory and growth accounting. It must also move beyond the first generation of endogenous growth models in which technological change is mechanically linked to the rate of growth of the capital stock. And ultimately, it must link together all the evidence that economists have on growth. Especially for questions posed at the aggregate level, information about exogeneity or causality is very scarce. The few natural experiments that can truly reveal something about the underlying causal mechanisms are so rare that they must all be considered if we are to make progress toward an understanding of aggregate economic growth.