Chapter 1

THE SOLOW GROWTH MODEL

1.1 Some Basic Facts about Economic Growth

Over the past few centuries, standards of living in industrialized countries have reached levels almost unimaginable to our ancestors. Although comparisons are difficult, the best available evidence suggests that average real incomes today in the United States and Western Europe are between 10 and 30 times larger than a century ago, and between 50 and 300 times larger than two centuries ago.[1]

Moreover, worldwide growth is far from constant. Growth has been rising over most of modern history. Average growth rates in the industrialized countries were higher in the twentieth century than in the nineteenth, and higher in the nineteenth than in the eighteenth. Further, average incomes on the eve of the Industrial Revolution even in the wealthiest countries were not dramatically above subsistence levels; this tells us that average growth over the millennia before the Industrial Revolution must have been very, very low.

One important exception to this general pattern of increasing growth is the *productivity growth slowdown.* Average annual growth in output per person in the United States and other industrialized countries from the early 1970s to the mid-1990s was about a percentage point below its earlier level. The data since then suggest a rebound in productivity growth, at least in the United States. How long the rebound will last and how widespread it will be are not yet clear.

[1] Maddison (2003) reports and discusses basic data on average real incomes over modern history. Most of the uncertainty about the extent of long-term growth concerns the behavior not of nominal income, but of the price indexes needed to convert those figures into estimates of real income. Adjusting for quality changes and for the introduction of new goods is conceptually and practically difficult, and conventional price indexes do not make these adjustments well. See Nordhaus (1997) and Boskin, Dulberger, Gordon, Griliches, and Jorgenson (1998) for discussions of the issues involved and analyses of the biases in conventional price indexes.

There are also enormous differences in standards of living across parts of the world. Average real incomes in such countries as the United States, Germany, and Japan appear to exceed those in such countries as Bangladesh and Kenya by a factor of about 20.[2] As with worldwide growth, cross-country income differences are not immutable. Growth in individual countries often differs considerably from average worldwide growth; that is, there are often large changes in countries' relative incomes.

The most striking examples of large changes in relative incomes are *growth miracles* and *growth disasters.* Growth miracles are episodes where growth in a country far exceeds the world average over an extended period, with the result that the country moves rapidly up the world income distribution. Some prominent growth miracles are Japan from the end of World War II to around 1990 and the newly industrializing countries (NICs) of East Asia—South Korea, Taiwan, Singapore, and Hong Kong—starting around 1960. Average incomes in the NICs, for example, have grown at an average annual rate of over 5 percent since 1960. As a result, their average incomes relative to that of the United States have more than tripled.

Growth disasters are episodes where a country's growth falls far short of the world average. Two very different examples of growth disasters are Argentina and many of the countries of sub-Saharan Africa. In 1900, Argentina's average income was only slightly behind those of the world's leaders, and it appeared poised to become a major industrialized country. But its growth performance over most of the twentieth century was dismal, and it is now near the middle of the world income distribution. Sub-Saharan African countries such as Chad, Ghana, and Mozambique have been extremely poor throughout their histories and have been unable to obtain any sustained growth in average incomes. As a result, their average incomes have remained close to subsistence levels while average world income has been rising steadily.

Other countries exhibit more complicated growth patterns. Cote d'Ivoire was held up as the growth model for Africa through the 1970s. From 1960 to 1978, real income per person grew at an average annual rate of 3.5 percent. But in the next decade, average income fell by a third. To take another example, average growth in Mexico was extremely high in the 1960s and 1970s, negative in most of the 1980s, and again very high—with a brief but severe interruption in the mid-1990s—since then.

Over the whole of the modern era, cross-country income differences have widened on average. The fact that average incomes in the richest countries at the beginning of the Industrial Revolution were not far above subsistence means that the overall dispersion of average incomes across different parts of the world must have been much smaller than it is today (Pritchett, 1997). Over the past few decades, however, there has been no strong tendency either toward continued divergence or toward convergence.

[2] Comparisons of real incomes across countries are far from straightforward, but are much easier than comparisons over extended periods of time. The basic source for cross-country data on real income is the Penn World Tables. Documentation of these data and the most recent figures are available at the National Bureau of Economic Research's web site, http://www.nber.org.

The implications of the vast differences in standards of living over time and across countries for human welfare are enormous. The differences are associated with large differences in nutrition, literacy, infant mortality, life expectancy, and other direct measures of well-being. And the welfare consequences of long-run growth swamp any possible effects of the short-run fluctuations that macroeconomics traditionally focuses on. During an average recession in the United States, for example, real income per person falls by a few percent relative to its usual path. In contrast, the productivity growth slowdown reduced real income per person in the United States by about 25 percent relative to what it otherwise would have been. Other examples are even more startling. If real income per person in Bangladesh continues to grow at its postwar average rate of 1.1 percent, it will take well over 200 years for it to reach the current U.S. level. If Bangladesh achieves 3 percent growth, the time will be reduced to 100 years. And if it achieves 5 percent growth, as the NICs have done, the process will take only 60 years. To quote Robert Lucas (1988), "Once one starts to think about [economic growth], it is hard to think about anything else."

The first three chapters of this book are therefore devoted to economic growth. We will investigate several models of growth. Although we will examine the models' mechanics in considerable detail, our goal is to learn what insights they offer concerning worldwide growth and income differences across countries. Indeed, the ultimate objective of research on economic growth is to determine whether there are possibilities for raising overall growth or bringing standards of living in poor countries closer to those in the world leaders.

This chapter focuses on the model that economists have traditionally used to study these issues, the Solow growth model.[3] The Solow model is the starting point for almost all analyses of growth. Even models that depart fundamentally from Solow's are often best understood through comparison with the Solow model. Thus understanding the model is essential to understanding theories of growth.

The principal conclusion of the Solow model is that the accumulation of physical capital cannot account for either the vast growth over time in output per person or the vast geographic differences in output per person. Specifically, suppose that capital accumulation affects output through the conventional channel that capital makes a direct contribution to production, for which it is paid its marginal product. Then the Solow model implies that the differences in real incomes that we are trying to understand are far

[3] The Solow model (which is sometimes known as the Solow–Swan model) was developed by Robert Solow (Solow, 1956) and T. W. Swan (Swan, 1956).

too large to be accounted for by differences in capital inputs. The model treats other potential sources of differences in real incomes as either exogenous and thus not explained by the model (in the case of technological progress, for example) or absent altogether (in the case of positive externalities from capital, for example). Thus to address the central questions of growth theory, we must move beyond the Solow model.

Chapters 2 and 3 therefore extend and modify the Solow model. Chapter 2 investigates the determinants of saving and investment. The Solow model has no optimization in it; it simply takes the saving rate as exogenous and constant. Chapter 2 presents two models that make saving endogenous and potentially time-varying. In the first, saving and consumption decisions are made by a fixed set of infinitely lived households; in the second, the decisions are made by overlapping generations of households with finite horizons.

Relaxing the Solow model's assumption of a constant saving rate has three advantages. First, and most important for studying growth, it demonstrates that the Solow model's conclusions about the central questions of growth theory do not hinge on its assumption of a fixed saving rate. Second, it allows us to consider welfare issues. A model that directly specifies relations among aggregate variables provides no way of judging whether some outcomes are better or worse than others: without individuals in the model, we cannot say whether different outcomes make individuals better or worse off. The infinite-horizon and overlapping-generations models are built up from the behavior of individuals, and therefore can be used to discuss welfare issues. Third, infinite-horizon and overlapping-generations models are used to study many issues in economics other than economic growth; thus they are valuable tools.

Chapter 3 investigates more fundamental departures from the Solow model. Its models, in contrast to Chapter 2's, provide different answers than the Solow model to the central questions of growth theory. The first part of the chapter departs from the Solow model's treatment of technological progress as exogenous; it assumes instead that it is the result of the allocation of resources to the creation of new technologies. We will investigate the implications of such *endogenous technological progress* for economic growth and the determinants of the allocation of resources to innovative activities.

The main conclusion of this analysis is that endogenous technological progress is almost surely central to worldwide growth but probably has little to do with cross-country income differences. The second part of Chapter 3 therefore focuses specifically on those differences. We will find that understanding those differences requires considering two new factors: differences in human as well as physical capital, and differences in productivity not stemming from differences in technology. This material explores both how those factors can help us understand the enormous differences

in average incomes across countries and potential sources of differences in those factors.

We now turn to the Solow model.

1.2 Assumptions

Inputs and Output

The Solow model focuses on four variables: output (Y), capital (K), labor (L), and "knowledge" or the "effectiveness of labor" (A). At any time, the economy has some amounts of capital, labor, and knowledge, and these are combined to produce output. The production function takes the form

$$Y(t) = F(K(t), A(t)L(t)), \tag{1.1}$$

where t denotes time.

Notice that time does not enter the production function directly, but only through K, L, and A. That is, output changes over time only if the inputs to production change. In particular, the amount of output obtained from given quantities of capital and labor rises over time—there is technological progress—only if the amount of knowledge increases.

Notice also that A and L enter multiplicatively. AL is referred to as *effective labor,* and technological progress that enters in this fashion is known as *labor-augmenting* or *Harrod-neutral.*[4] This way of specifying how A enters, together with the other assumptions of the model, will imply that the ratio of capital to output, K/Y, eventually settles down. In practice, capital-output ratios do not show any clear upward or downward trend over extended periods. In addition, building the model so that the ratio is eventually constant makes the analysis much simpler. Assuming that A multiplies L is therefore very convenient.

The central assumptions of the Solow model concern the properties of the production function and the evolution of the three inputs into production (capital, labor, and knowledge) over time. We discuss each in turn.

Assumptions Concerning the Production Function

The model's critical assumption concerning the production function is that it has constant returns to scale in its two arguments, capital and effective labor. That is, doubling the quantities of capital and effective labor (for example, by doubling K and L with A held fixed) doubles the amount produced.

[4] If knowledge enters in the form $Y = F(AK, L)$, technological progress is *capital-augmenting.* If it enters in the form $Y = AF(K, L)$, technological progress is *Hicks-neutral.*

More generally, multiplying both arguments by any nonnegative constant c causes output to change by the same factor:

$$F(cK, cAL) = cF(K, AL) \quad \text{for all } c \geq 0. \tag{1.2}$$

The assumption of constant returns can be thought of as a combination of two separate assumptions. The first is that the economy is big enough that the gains from specialization have been exhausted. In a very small economy, there are probably enough possibilities for further specialization that doubling the amounts of capital and labor more than doubles output. The Solow model assumes, however, that the economy is sufficiently large that, if capital and labor double, the new inputs are used in essentially the same way as the existing inputs, and thus that output doubles.

The second assumption is that inputs other than capital, labor, and knowledge are relatively unimportant. In particular, the model neglects land and other natural resources. If natural resources are important, doubling capital and labor could less than double output. In practice, however, as Section 1.8 describes, the availability of natural resources does not appear to be a major constraint on growth. Assuming constant returns to capital and labor alone therefore appears to be a reasonable approximation.

The assumption of constant returns allows us to work with the production function in *intensive form.* Setting $c = 1/AL$ in equation (1.2) yields

$$F\left(\frac{K}{AL}, 1\right) = \frac{1}{AL}F(K, AL). \tag{1.3}$$

Here K/AL is the amount of capital per unit of effective labor, and $F(K, AL)/AL$ is Y/AL, output per unit of effective labor. Define $k = K/AL$, $y = Y/AL$, and $f(k) = F(k, 1)$. Then we can rewrite (1.3) as

$$y = f(k). \tag{1.4}$$

That is, we can write output per unit of effective labor as a function of capital per unit of effective labor.

These new variables, k and y, are not of interest in their own right. Rather, they are tools for learning about the variables we are interested in. As we will see, the easiest way to analyze the model is to focus on the behavior of k rather than to consider directly the behavior of the two arguments of the production function, K and AL. For example, we will determine the behavior of output per worker, Y/L, by writing it as $A(Y/AL)$, or $Af(k)$, and determining the behavior of A and k.

To see the intuition behind (1.4), think of dividing the economy into AL small economies, each with 1 unit of effective labor and K/AL units of capital. Since the production function has constant returns, each of these small economies produces $1/AL$ as much as is produced in the large, undivided economy. Thus the amount of output per unit of effective labor depends only on the quantity of capital per unit of effective labor, and not on the overall size of the economy. This is expressed mathematically in equation (1.4).

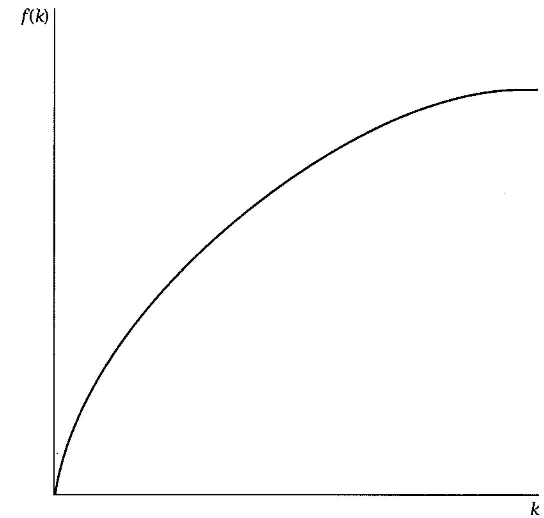

FIGURE 1.1 An example of a production function

The intensive-form production function, $f(k)$, is assumed to satisfy $f(0) = 0$, $f'(k) > 0$, $f''(k) < 0$.[5] Since $F(K, AL)$ equals $ALf(K/AL)$, it follows that the marginal product of capital, $\partial F(K, AL)/\partial K$, equals $ALf'(K/AL)(1/AL)$, which is just $f'(k)$. Thus the assumptions that $f'(k)$ is positive and $f''(k)$ is negative imply that the marginal product of capital is positive, but that it declines as capital (per unit of effective labor) rises. In addition, $f(\bullet)$ is assumed to satisfy the *Inada conditions* (Inada, 1964): $\lim_{k\to 0} f'(k) = \infty$, $\lim_{k\to\infty} f'(k) = 0$. These conditions (which are stronger than needed for the model's central results) state that the marginal product of capital is very large when the capital stock is sufficiently small and that it becomes very small as the capital stock becomes large; their role is to ensure that the path of the economy does not diverge. A production function satisfying $f'(\bullet) > 0$, $f''(\bullet) < 0$, and the Inada conditions is shown in Figure 1.1.

A specific example of a production function is the Cobb–Douglas function,

$$F(K, AL) = K^{\alpha}(AL)^{1-\alpha}, \qquad 0 < \alpha < 1. \tag{1.5}$$

This production function is easy to analyze, and it appears to be a good first approximation to actual production functions. As a result, it is very useful.

[5] The notation $f'(\bullet)$ denotes the first derivative of $f(\bullet)$, and $f''(\bullet)$ the second derivative.

It is easy to check that the Cobb–Douglas function has constant returns. Multiplying both inputs by c gives us

$$\begin{aligned} F(cK, cAL) &= (cK)^{\alpha}(cAL)^{1-\alpha} \\ &= c^{\alpha}c^{1-\alpha}K^{\alpha}(AL)^{1-\alpha} \\ &= cF(K, AL). \end{aligned} \tag{1.6}$$

To find the intensive form of the production function, divide both inputs by AL; this yields

$$\begin{aligned} f(k) &\equiv F\left(\frac{K}{AL}, 1\right) \\ &= \left(\frac{K}{AL}\right)^{\alpha} \\ &= k^{\alpha}. \end{aligned} \tag{1.7}$$

Equation (1.7) implies that $f'(k) = \alpha k^{\alpha-1}$. It is straightforward to check that this expression is positive, that it approaches infinity as k approaches zero, and that it approaches zero as k approaches infinity. Finally, $f''(k) = -(1-\alpha)\alpha k^{\alpha-2}$, which is negative.[6]

The Evolution of the Inputs into Production

The remaining assumptions of the model concern how the stocks of labor, knowledge, and capital change over time. The model is set in continuous time; that is, the variables of the model are defined at every point in time.[7]

The initial levels of capital, labor, and knowledge are taken as given. Labor and knowledge grow at constant rates:

$$\dot{L}(t) = nL(t), \tag{1.8}$$

$$\dot{A}(t) = gA(t), \tag{1.9}$$

where n and g are exogenous parameters and where a dot over a variable denotes a derivative with respect to time (that is, $\dot{X}(t)$ is shorthand for $dX(t)/dt$).

[6] Note that with Cobb–Douglas production, labor-augmenting, capital-augmenting, and Hicks-neutral technological progress (see n. 4) are all essentially the same. For example, to rewrite (1.5) so that technological progress is Hicks-neutral, simply define $\tilde{A} = A^{1-\alpha}$; then $Y = \tilde{A}(K^{\alpha}L^{1-\alpha})$.

[7] The alternative is discrete time, where the variables are defined only at specific dates (usually $t = 0, 1, 2, \ldots$). The choice between continuous and discrete time is usually based on convenience. For example, the Solow model has essentially the same implications in discrete as in continuous time, but is easier to analyze in continuous time.

The *growth rate* of a variable refers to its proportional rate of change. That is, *the growth rate of X* refers to the quantity $\dot{X}(t)/X(t)$. Thus equation (1.8) implies that the growth rate of L is constant and equal to n, and (1.9) implies that A's growth rate is constant and equal to g.

A key fact about growth rates is that the growth rate of a variable equals the rate of change of its natural log. That is, $\dot{X}(t)/X(t)$ equals $d\ln X(t)/dt$. To see this, note that since $\ln X$ is a function of X and X is a function of t, we can use the chain rule to write

$$\begin{aligned} \frac{d\ln X(t)}{dt} &= \frac{d\ln X(t)}{dX(t)}\frac{dX(t)}{dt} \\ &= \frac{1}{X(t)}\dot{X}(t). \end{aligned} \tag{1.10}$$

Applying the result that a variable's growth rate equals the rate of change of its log to (1.8) and (1.9) tells us that the rates of change of the logs of L and A are constant and that they equal n and g, respectively. Thus,

$$\ln L(t) = [\ln L(0)] + nt, \tag{1.11}$$

$$\ln A(t) = [\ln A(0)] + gt, \tag{1.12}$$

where $L(0)$ and $A(0)$ are the values of L and A at time 0. Exponentiating both sides of these equations gives us

$$L(t) = L(0)e^{nt}, \tag{1.13}$$

$$A(t) = A(0)e^{gt}. \tag{1.14}$$

Thus, our assumption is that L and A each grow exponentially.[8]

Output is divided between consumption and investment. The fraction of output devoted to investment, s, is exogenous and constant. One unit of output devoted to investment yields one unit of new capital. In addition, existing capital depreciates at rate δ. Thus

$$\dot{K}(t) = sY(t) - \delta K(t). \tag{1.15}$$

Although no restrictions are placed on n, g, and δ individually, their sum is assumed to be positive. This completes the description of the model.

Since this is the first model (of many!) we will encounter, this is a good place for a general comment about modeling. The Solow model is grossly simplified in a host of ways. To give just a few examples, there is only a single good; government is absent; fluctuations in employment are ignored; production is described by an aggregate production function with just three inputs; and the rates of saving, depreciation, population growth, and technological progress are constant. It is natural to think of these features of the model as defects: the model omits many obvious features of the world,

[8] See Problems 1.1 and 1.2 for more on basic properties of growth rates.

and surely some of those features are important to growth. But the purpose of a model is not to be realistic. After all, we already possess a model that is completely realistic—the world itself. The problem with that "model" is that it is too complicated to understand. A model's purpose is to provide insights about particular features of the world. If a simplifying assumption causes a model to give incorrect answers *to the questions it is being used to address,* then that lack of realism may be a defect. (Even then, the simplification—by showing clearly the consequences of those features of the world in an idealized setting—may be a useful reference point.) If the simplification does not cause the model to provide incorrect answers to the questions it is being used to address, however, then the lack of realism is a virtue: by isolating the effect of interest more clearly, the simplification makes it easier to understand.

1.3 The Dynamics of the Model

We want to determine the behavior of the economy we have just described. The evolution of two of the three inputs into production, labor and knowledge, is exogenous. Thus to characterize the behavior of the economy, we must analyze the behavior of the third input, capital.

The Dynamics of k

Because the economy may be growing over time, it turns out to be much easier to focus on the capital stock per unit of effective labor, k, than on the unadjusted capital stock, K. Since $k = K/AL$, we can use the chain rule to find

$$\begin{aligned}\dot{k}(t) &= \frac{\dot{K}(t)}{A(t)L(t)} - \frac{K(t)}{[A(t)L(t)]^2}[A(t)\dot{L}(t) + L(t)\dot{A}(t)] \\ &= \frac{\dot{K}(t)}{A(t)L(t)} - \frac{K(t)}{A(t)L(t)}\frac{\dot{L}(t)}{L(t)} - \frac{K(t)}{A(t)L(t)}\frac{\dot{A}(t)}{A(t)}.\end{aligned} \tag{1.16}$$

K/AL is simply k. From (1.8) and (1.9), $\dot{L}/L$ and $\dot{A}/A$ are n and g, respectively. $\dot{K}$ is given by (1.15). Substituting these facts into (1.16) yields

$$\begin{aligned}\dot{k}(t) &= \frac{sY(t) - \delta K(t)}{A(t)L(t)} - k(t)n - k(t)g \\ &= s\frac{Y(t)}{A(t)L(t)} - \delta k(t) - nk(t) - gk(t).\end{aligned} \tag{1.17}$$

Finally, using the fact that Y/AL is given by $f(k)$, we have

$$\dot{k}(t) = sf(k(t)) - (n + g + \delta)k(t). \tag{1.18}$$

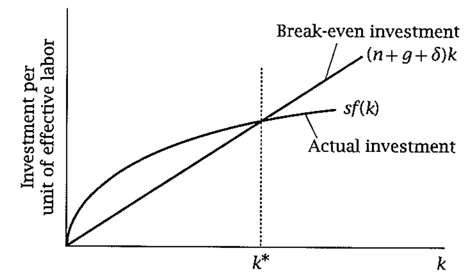

FIGURE 1.2 Actual and break-even investment

Equation (1.18) is the key equation of the Solow model. It states that the rate of change of the capital stock per unit of effective labor is the difference between two terms. The first, $sf(k)$, is actual investment per unit of effective labor: output per unit of effective labor is $f(k)$, and the fraction of that output that is invested is s. The second term, $(n + g + \delta)k$, is *break-even investment,* the amount of investment that must be done just to keep k at its existing level. There are two reasons that some investment is needed to prevent k from falling. First, existing capital is depreciating; this capital must be replaced to keep the capital stock from falling. This is the δk term in (1.18). Second, the quantity of effective labor is growing. Thus doing enough investment to keep the capital stock (K) constant is not enough to keep the capital stock per unit of effective labor (k) constant. Instead, since the quantity of effective labor is growing at rate $n + g$, the capital stock must grow at rate $n + g$ to hold k steady.[9] This is the $(n + g)k$ term in (1.18).

When actual investment per unit of effective labor exceeds the investment needed to break even, k is rising. When actual investment falls short of break-even investment, k is falling. And when the two are equal, k is constant.

Figure 1.2 plots the two terms of the expression for $\dot{k}$ as functions of k. Break-even investment, $(n + g + \delta)k$, is proportional to k. Actual investment, $sf(k)$, is a constant times output per unit of effective labor.

Since $f(0) = 0$, actual investment and break-even investment are equal at $k = 0$. The Inada conditions imply that at $k = 0$, $f'(k)$ is large, and thus that the $sf(k)$ line is steeper than the $(n + g + \delta)k$ line. Thus for small values of k, actual investment is larger than break-even investment. The Inada conditions also imply that $f'(k)$ falls toward zero as k becomes large. At some

[9] The fact that the growth rate of the quantity of effective labor, AL, equals $n + g$ is an instance of the fact that the growth rate of the product of two variables equals the sum of their growth rates. See Problem 1.1.

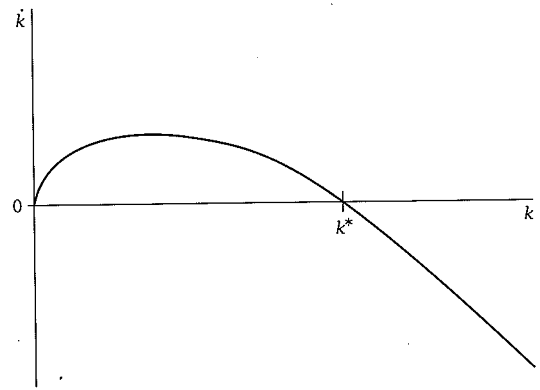

FIGURE 1.3 The phase diagram for k in the Solow model

point, the slope of the actual investment line falls below the slope of the break-even investment line. With the $sf(k)$ line flatter than the $(n+g+\delta)k$ line, the two must eventually cross. Finally, the fact that $f''(k) < 0$ implies that the two lines intersect only once for $k > 0$. We let k^* denote the value of k where actual investment and break-even investment are equal.

Figure 1.3 summarizes this information in the form of a *phase diagram*, which shows $\dot{k}$ as a function of k. If k is initially less than k^*, actual investment exceeds break-even investment, and so $\dot{k}$ is positive—that is, k is rising. If k exceeds k^*, $\dot{k}$ is negative. Finally, if k equals k^*, then $\dot{k}$ is zero. Thus, regardless of where k starts, it converges to k^*.[10]

The Balanced Growth Path

Since k converges to k^*, it is natural to ask how the variables of the model behave when k equals k^*. By assumption, labor and knowledge are growing at rates n and g, respectively. The capital stock, K, equals ALk; since k is constant at k^*, K is growing at rate $n+g$ (that is, $\dot{K}/K$ equals $n+g$). With both capital and effective labor growing at rate $n+g$, the assumption of constant returns implies that output, Y, is also growing at that rate. Finally, capital per worker, K/L, and output per worker, Y/L, are growing at rate g.

Thus the Solow model implies that, regardless of its starting point, the economy converges to a *balanced growth path*—a situation where each

[10] If k is initially zero, it remains there. We ignore this possibility in what follows.

variable of the model is growing at a constant rate. On the balanced growth path, the growth rate of output per worker is determined solely by the rate of technological progress.[11]

1.4 The Impact of a Change in the Saving Rate

The parameter of the Solow model that policy is most likely to affect is the saving rate. The division of the government's purchases between consumption and investment goods, the division of its revenues between taxes and borrowing, and its tax treatments of saving and investment are all likely to affect the fraction of output that is invested. Thus it is natural to investigate the effects of a change in the saving rate.

For concreteness, we will consider a Solow economy that is on a balanced growth path, and suppose that there is a permanent increase in s. In addition to demonstrating the model's implications concerning the role of saving, this experiment will illustrate the model's properties when the economy is not on a balanced growth path.

The Impact on Output

The increase in s shifts the actual investment line upward, and so k^* rises. This is shown in Figure 1.4. But k does not immediately jump to the new value of k^*. Initially, k is equal to the old value of k^*. At this level, actual investment now exceeds break-even investment—more resources are being devoted to investment than are needed to hold k constant—and so $\dot{k}$ is positive. Thus k begins to rise. It continues to rise until it reaches the new value of k^*, at which point it remains constant.

These results are summarized in the first three panels of Figure 1.5. t_0 denotes the time of the increase in the saving rate. By assumption, s jumps up at time t_0 and remains constant thereafter. Since the jump in s causes actual investment to exceed break-even investment by a strictly positive amount,

[11] The broad behavior of the U.S. economy and many other major industrialized economies over the last century or more is described reasonably well by the balanced growth path of the Solow model. The growth rates of labor, capital, and output have each been roughly constant. The growth rates of output and capital have been about equal (so that the capital-output ratio has been approximately constant) and have been larger than the growth rate of labor (so that output per worker and capital per worker have been rising). This is often taken as evidence that it is reasonable to think of these economies as Solow-model economies on their balanced growth paths. Jones (2002a) shows, however, that the underlying determinants of the level of income on the balanced growth path have in fact been far from constant in these economies, and thus that the resemblance between these economies and the balanced growth path of the Solow model is misleading. We return to this issue in Section 3.3.

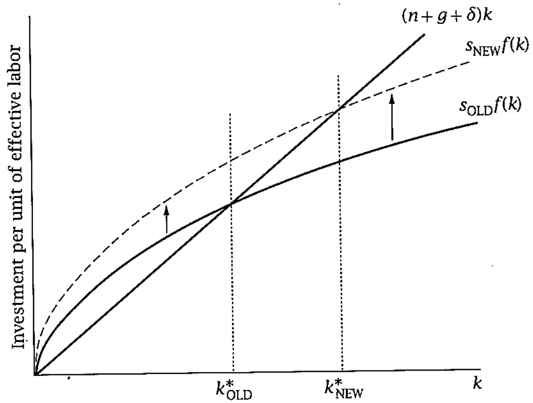

FIGURE 1.4 **The effects of an increase in the saving rate on investment**

$\dot{k}$ jumps from zero to a strictly positive amount. k rises gradually from the old value of k^* to the new value, and $\dot{k}$ falls gradually back to zero.[12]

We are likely to be particularly interested in the behavior of output per worker, Y/L. Y/L equals $Af(k)$. When k is constant, Y/L grows at rate g, the growth rate of A. When k is increasing, Y/L grows both because A is increasing and because k is increasing. Thus its growth rate exceeds g. When k reaches the new value of k^*, however, again only the growth of A contributes to the growth of Y/L, and so the growth rate of Y/L returns to g. Thus a *permanent* increase in the saving rate produces a *temporary* increase in the growth rate of output per worker: k is rising for a time, but eventually it increases to the point where the additional saving is devoted entirely to maintaining the higher level of k.

The fourth and fifth panels of Figure 1.5 show how output per worker responds to the rise in the saving rate. The *growth rate* of output per worker, which is initially g, jumps upward at t_0 and then gradually returns to its initial level. Thus output per worker begins to rise above the path it was on and gradually settles into a higher path parallel to the first.[13]

In sum, a change in the saving rate has a *level effect* but not a *growth effect:* it changes the economy's balanced growth path, and thus the level of

[12] For a sufficiently large rise in the saving rate, $\dot{k}$ rises for a while after t_0 before starting to fall back to zero.

[13] Because the growth rate of a variable equals the derivative with respect to time of its log, graphs in logs are often much easier to interpret than graphs in levels. For example, if a variable's growth rate is constant, the graph of its log as a function of time is a straight line. This is why Figure 1.5 shows the log of output per worker rather than its level.

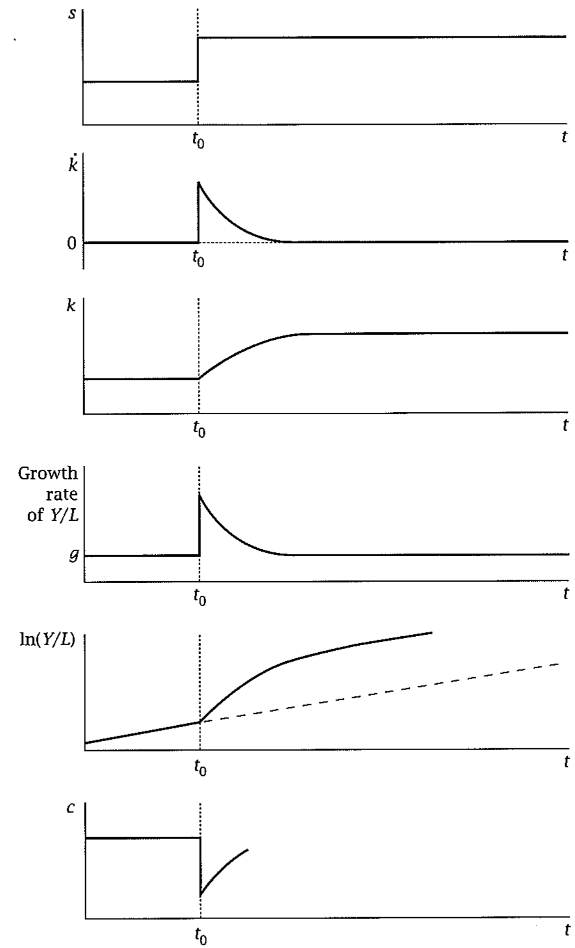

FIGURE 1.5 **The effects of an increase in the saving rate**

output per worker at any point in time, but it does not affect the growth rate of output per worker on the balanced growth path. Indeed, in the Solow model only changes in the rate of technological progress have growth effects; all other changes have only level effects.

The Impact on Consumption

If we were to introduce households into the model, their welfare would depend not on output but on consumption: investment is simply an input into production in the future. Thus for many purposes we are likely to be more interested in the behavior of consumption than in the behavior of output.

Consumption per unit of effective labor equals output per unit of effective labor, $f(k)$, times the fraction of that output that is consumed, $1 - s$. Thus, since s changes discontinuously at t_0 and k does not, initially consumption per unit of effective labor jumps downward. Consumption then rises gradually as k rises and s remains at its higher level. This is shown in the last panel of Figure 1.5.

Whether consumption eventually exceeds its level before the rise in s is not immediately clear. Let c^* denote consumption per unit of effective labor on the balanced growth path. c^* equals output per unit of effective labor, $f(k^*)$, minus investment per unit of effective labor, $sf(k^*)$. On the balanced growth path, actual investment equals break-even investment, $(n+g+\delta)k^*$. Thus,

$$c^* = f(k^*) - (n+g+\delta)k^*. \tag{1.19}$$

k^* is determined by s and the other parameters of the model, n, g, and δ; we can therefore write $k^* = k^*(s, n, g, \delta)$. Thus (1.19) implies

$$\frac{\partial c^*}{\partial s} = [f'(k^*(s,n,g,\delta)) - (n+g+\delta)]\frac{\partial k^*(s,n,g,\delta)}{\partial s}. \tag{1.20}$$

We know that the increase in s raises k^*. Thus whether the increase raises or lowers consumption in the long run depends on whether $f'(k^*)$—the marginal product of capital—is more or less than $n+g+\delta$. Intuitively, when k rises, investment (per unit of effective labor) must rise by $n+g+\delta$ times the change in k for the increase to be sustained. If $f'(k^*)$ is less than $n+g+\delta$, then the additional output from the increased capital is not enough to maintain the capital stock at its higher level. In this case, consumption must fall to maintain the higher capital stock. If $f'(k^*)$ exceeds $n+g+\delta$, on the other hand, there is more than enough additional output to maintain k at its higher level, and so consumption rises.

$f'(k^*)$ can be either smaller or larger than $n+g+\delta$. This is shown in Figure 1.6. The figure shows not only $(n+g+\delta)k$ and $sf(k)$, but also $f(k)$. Since consumption on the balanced growth path equals output less break-even investment (see [1.19]), c^* is the distance between $f(k)$ and $(n+g+\delta)k$ at $k = k^*$. The figure shows the determinants of c^* for three different values of s (and hence three different values of k^*). In the top panel, s is high, and so k^* is high and $f'(k^*)$ is less than $n+g+\delta$. As a result, an increase in the

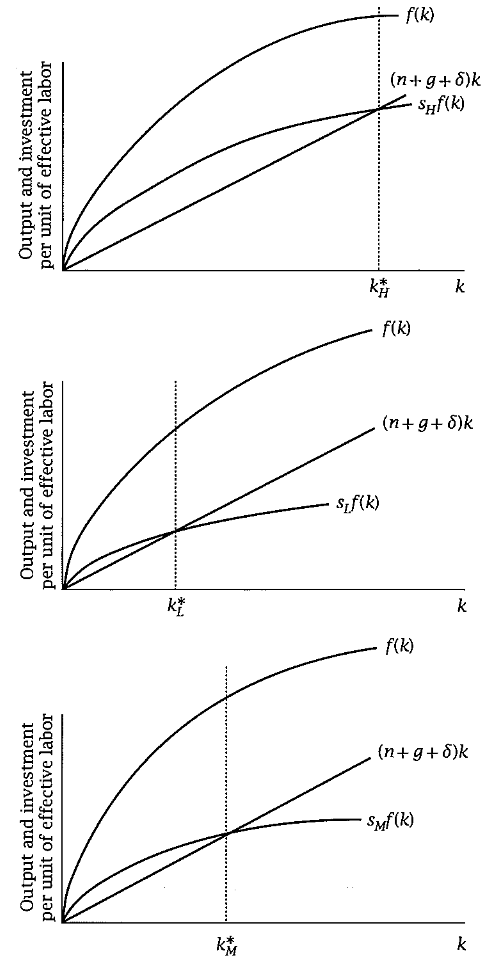

FIGURE 1.6 Output, investment, and consumption on the balanced growth path

saving rate lowers consumption even when the economy has reached its new balanced growth path. In the middle panel, s is low, k^* is low, $f'(k^*)$ is greater than $n+g+\delta$, and an increase in s raises consumption in the long run.

Finally, in the bottom panel, s is at the level that causes $f'(k^*)$ to just equal $n+g+\delta$—that is, the $f(k)$ and $(n+g+\delta)k$ loci are parallel at $k = k^*$. In this case, a marginal change in s has no effect on consumption in the long run, and consumption is at its maximum possible level among balanced growth paths. This value of k^* is known as the *golden-rule* level of the capital stock. We will discuss the golden-rule capital stock further in Chapter 2. Among the questions we will address are whether the golden-rule capital stock is in fact desirable and whether there are situations in which a decentralized economy with endogenous saving converges to that capital stock. Of course, in the Solow model, where saving is exogenous, there is no more reason to expect the capital stock on the balanced growth path to equal the golden-rule level than there is to expect it to equal any other possible value.

1.5 Quantitative Implications

We are often interested not just in a model's qualitative implications, but in its quantitative predictions. If, for example, the impact of a moderate increase in saving on growth remains large after several centuries, the result that the impact is temporary is of limited interest.

For most models, including this one, obtaining exact quantitative results requires specifying functional forms and values of the parameters; it often also requires analyzing the model numerically. But in many cases, it is possible to learn a great deal by considering approximations around the long-run equilibrium. That is the approach we take here.

The Effect on Output in the Long Run

The long-run effect of a rise in saving on output is given by

$$\frac{\partial y^*}{\partial s} = f'(k^*)\frac{\partial k^*(s,n,g,\delta)}{\partial s}, \tag{1.21}$$

where $y^* = f(k^*)$ is the level of output per unit of effective labor on the balanced growth path. Thus to find $\partial y^*/\partial s$, we need to find $\partial k^*/\partial s$. To do this, note that k^* is defined by the condition that $\dot{k} = 0$. Thus k^* satisfies

$$sf(k^*(s,n,g,\delta)) = (n+g+\delta)k^*(s,n,g,\delta). \tag{1.22}$$

Equation (1.22) holds for all values of s (and of n, g, and δ). Thus the derivatives of the two sides with respect to s are equal:[14]

$$sf'(k^*)\frac{\partial k^*}{\partial s} + f(k^*) = (n+g+\delta)\frac{\partial k^*}{\partial s}, \tag{1.23}$$

where the arguments of k^* are omitted for simplicity. This can be rearranged to obtain[15]

$$\frac{\partial k^*}{\partial s} = \frac{f(k^*)}{(n+g+\delta) - sf'(k^*)}. \tag{1.24}$$

Substituting (1.24) into (1.21) yields

$$\frac{\partial y^*}{\partial s} = \frac{f'(k^*)f(k^*)}{(n+g+\delta) - sf'(k^*)}. \tag{1.25}$$

Two changes help in interpreting this expression. The first is to convert it to an elasticity by multiplying both sides by s/y^*. The second is to use the fact that $sf(k^*) = (n+g+\delta)k^*$ to substitute for s. Making these changes gives us

$$\begin{aligned}\frac{s}{y^*}\frac{\partial y^*}{\partial s} &= \frac{s}{f(k^*)}\frac{f'(k^*)f(k^*)}{(n+g+\delta) - sf'(k^*)} \\ &= \frac{(n+g+\delta)k^*f'(k^*)}{f(k^*)[(n+g+\delta) - (n+g+\delta)k^*f'(k^*)/f(k^*)]} \\ &= \frac{k^*f'(k^*)/f(k^*)}{1 - [k^*f'(k^*)/f(k^*)]}.\end{aligned} \tag{1.26}$$

$k^*f'(k^*)/f(k^*)$ is the elasticity of output with respect to capital at $k = k^*$. Denoting this by $\alpha_K(k^*)$, we have

$$\frac{s}{y^*}\frac{\partial y^*}{\partial s} = \frac{\alpha_K(k^*)}{1-\alpha_K(k^*)}. \tag{1.27}$$

If markets are competitive and there are no externalities, capital earns its marginal product. Since output equals $ALf(k)$ and k equals K/AL, the marginal product of capital, $\partial Y/\partial K$, is $ALf'(k)[1/(AL)]$, or just $f'(k)$. Thus if

[14] This technique is known as *implicit differentiation*. Even though (1.22) does not explicitly give k^* as a function of s, n, g, and δ, it still determines how k^* depends on those variables. We can therefore differentiate the equation with respect to s and solve for $\partial k^*/\partial s$.

[15] We saw in the previous section that an increase in s raises k^*. To check that this is also implied by equation (1.24), note that $n+g+\delta$ is the slope of the break-even investment line and that $sf'(k^*)$ is the slope of the actual investment line at k^*. Since the break-even investment line is steeper than the actual investment line at k^* (see Figure 1.2), it follows that the denominator of (1.24) is positive and thus that $\partial k^*/\partial s > 0$.

capital earns its marginal product, the total amount earned by capital (per unit of effective labor) on the balanced growth path is $k^*f'(k^*)$. The share of total income that goes to capital on the balanced growth path is then $k^*f'(k^*)/f(k^*)$, or $\alpha_K(k^*)$.

In most countries, the share of income paid to capital is about one-third. If we use this as an estimate of $\alpha_K(k^*)$, it follows that the elasticity of output with respect to the saving rate in the long run is about one-half. Thus, for example, a 10 percent increase in the saving rate (from 20 percent of output to 22 percent, for instance) raises output per worker in the long run by about 5 percent relative to the path it would have followed. Even a 50 percent increase in s raises y^* only by about 22 percent. Thus significant changes in saving have only moderate effects on the level of output on the balanced growth path.

Intuitively, a small value of $\alpha_K(k^*)$ makes the impact of saving on output low for two reasons. First, it implies that the actual investment curve, $sf(k)$, bends fairly sharply. As a result, an upward shift of the curve moves its intersection with the break-even investment line relatively little. Thus the impact of a change in s on k^* is small. Second, a low value of $\alpha_K(k^*)$ means that the impact of a change in k^* on y^* is small.

The Speed of Convergence

In practice, we are interested not only in the eventual effects of some change (such as a change in the saving rate), but also in how rapidly those effects occur. Again, we can use approximations around the long-run equilibrium to address this issue.

For simplicity, we focus on the behavior of k rather than y. Our goal is thus to determine how rapidly k approaches k^*. We know that $\dot{k}$ is determined by k: recall that the key equation of the model is $\dot{k} = sf(k) - (n+g+\delta)k$ (see [1.18]). Thus we can write $\dot{k} = \dot{k}(k)$. When k equals k^*, $\dot{k}$ is zero. A first-order Taylor-series approximation of $\dot{k}(k)$ around $k = k^*$ therefore yields

$$\dot{k} \simeq \left[\left. \frac{\partial \dot{k}(k)}{\partial k} \right|_{k=k^*} \right] (k - k^*). \tag{1.28}$$

That is, $\dot{k}$ is approximately equal to the product of the difference between k and k^* and the derivative of $\dot{k}$ with respect to k at $k = k^*$.

Let λ denote $-\partial \dot{k}(k)/\partial k|_{k=k^*}$. With this definition, (1.28) becomes

$$\dot{k}(t) \simeq -\lambda[k(t) - k^*]. \tag{1.29}$$

Since $\dot{k}$ is positive when k is slightly below k^* and negative when it is slightly above, $\partial \dot{k}(k)/\partial k|_{k=k^*}$ is negative. Equivalently, λ is positive.

Equation (1.29) implies that in the vicinity of the balanced growth path, k moves toward k^* at a speed approximately proportional to its distance from k^*. That is, the growth rate of $k(t) - k^*$ is approximately constant and equal to $-\lambda$. This implies

$$k(t) \simeq k^* + e^{-\lambda t}[k(0) - k^*], \tag{1.30}$$

where $k(0)$ is the initial value of k. Note that (1.30) follows just from the facts that the system is stable (that is, that k converges to k^*) and that we are linearizing the equation for $\dot{k}$ around $k = k^*$.

It remains to find λ; this is where the specifics of the model enter the analysis. Differentiating expression (1.18) for $\dot{k}$ with respect to k and evaluating the resulting expression at $k = k^*$ yields

$$\begin{aligned} \lambda \equiv - \left. \frac{\partial \dot{k}(k)}{\partial k} \right|_{k=k^*} &= -[sf'(k^*) - (n+g+\delta)] \\ &= (n+g+\delta) - sf'(k^*) \\ &= (n+g+\delta) - \frac{(n+g+\delta)k^*f'(k^*)}{f(k^*)} \\ &= [1 - \alpha_K(k^*)](n+g+\delta), \end{aligned} \tag{1.31}$$

where the third line again uses the fact that $sf(k^*) = (n+g+\delta)k^*$ to substitute for s, and where the last line uses the definition of α_K. Thus, k converges to its balanced-growth-path value at rate $[1 - \alpha_K(k^*)](n+g+\delta)$. In addition, one can show that y approaches y^* at the same rate that k approaches k^*. That is, $y(t) - y^* \simeq e^{-\lambda t}[y(0) - y^*]$.[16]

We can calibrate (1.31) to see how quickly actual economies are likely to approach their balanced growth paths. Typically, $n+g+\delta$ is about 6 percent per year (this would arise, for example, with 1 to 2 percent population growth, 1 to 2 percent growth in output per worker, and 3 to 4 percent depreciation). If capital's share is roughly one-third, $(1 - \alpha_K)(n+g+\delta)$ is thus roughly 4 percent. Therefore k and y move 4 percent of the remaining distance toward k^* and y^* each year, and take approximately 17 years to get halfway to their balanced-growth-path values.[17] Thus in our example of

[16] See Problem 1.11.

[17] The time it takes for a variable (in this case, $y - y^*$) with a constant negative growth rate to fall in half is approximately equal to 70 divided by its growth rate in percent. (Similarly, the doubling time of a variable with positive growth is 70 divided by the growth rate.) Thus in this case the *half-life* is roughly 70/(4%/year), or about 17 years. More exactly, the half-life, t^*, is the solution to $e^{-\lambda t^*} = 0.5$, where λ is the rate of decrease. Taking logs of both sides, $t^* = -\ln(0.5)/\lambda \simeq 0.69/\lambda$.

a 10 percent increase in the saving rate, output is 0.04(5%) = 0.2% above its previous path after 1 year; is 0.5(5%) = 2.5% above after 17 years; and asymptotically approaches 5 percent above the previous path. Thus not only is the overall impact of a substantial change in the saving rate modest, but it does not occur very quickly.[18]

1.6 The Solow Model and the Central Questions of Growth Theory

The Solow model identifies two possible sources of variation—either over time or across parts of the world—in output per worker: differences in capital per worker (K/L) and differences in the effectiveness of labor (A). We have seen, however, that only growth in the effectiveness of labor can lead to permanent growth in output per worker, and that for reasonable cases the impact of changes in capital per worker on output per worker is modest. As a result, only differences in the effectiveness of labor have any reasonable hope of accounting for the vast differences in wealth across time and space. Specifically, the central conclusion of the Solow model is that if the returns that capital commands in the market are a rough guide to its contributions to output, then variations in the accumulation of physical capital do not account for a significant part of either worldwide economic growth or cross-country income differences.

There are two ways to see that the Solow model implies that differences in capital accumulation cannot account for large differences in incomes, one direct and the other indirect. The direct approach is to consider the required differences in capital per worker. Suppose we want to account for a difference of a factor of X in output per worker between two economies on the basis of differences in capital per worker. If output per worker differs by a factor of X, the difference in log output per worker between the two economies is $\ln X$. Since the elasticity of output per worker with respect to capital per worker is α_K, log capital per worker must differ by $(\ln X)/\alpha_K$. That is, capital per worker differs by a factor of $e^{(\ln X)/\alpha_K}$, or X^{1/α_K}.

Output per worker in the major industrialized countries today is on the order of 10 times larger than it was 100 years ago, and 10 times larger than

[18] These results are derived from a Taylor-series approximation around the balanced growth path. Thus, formally, we can rely on them only in an arbitrarily small neighborhood around the balanced growth path. The question of whether Taylor-series approximations provide good guides for finite changes does not have a general answer. For the Solow model with conventional production functions, and for moderate changes in parameter values (such as those we have been considering), the Taylor-series approximations are generally quite reliable.

it is in poor countries today. Thus we would like to account for values of X in the vicinity of 10. Our analysis implies that doing this on the basis of differences in capital requires a difference of a factor of $10^{1/\alpha_K}$ in capital per worker. For $\alpha_K = \frac{1}{3}$, this is a factor of 1000. Even if capital's share is one-half, which is well above what data on capital income suggest, one still needs a difference of a factor of 100.

There is no evidence of such differences in capital stocks. Capital-output ratios are roughly constant over time. Thus the capital stock per worker in industrialized countries is roughly 10 times larger than it was 100 years ago, not 100 or 1000 times larger. Similarly, although capital-output ratios vary somewhat across countries, the variation is not great. For example, the capital-output ratio appears to be 2 to 3 times larger in industrialized countries than in poor countries; thus capital per worker is "only" about 20 to 30 times larger. In sum, differences in capital per worker are far smaller than those needed to account for the differences in output per worker that we are trying to understand.

The indirect way of seeing that the model cannot account for large variations in output per worker on the basis of differences in capital per worker is to notice that the required differences in capital imply enormous differences in the rate of return on capital (Lucas, 1990). If markets are competitive, the rate of return on capital equals its marginal product, $f'(k)$, minus depreciation, δ. Suppose that the production function is Cobb–Douglas (see equation [1.5]), which in intensive form is $f(k) = k^{\alpha}$. With this production function, the elasticity of output with respect to capital is simply α. The marginal product of capital is

$$\begin{aligned} f'(k) &= \alpha k^{\alpha-1} \\ &= \alpha y^{(\alpha-1)/\alpha}. \end{aligned} \tag{1.32}$$

Equation (1.32) implies that the elasticity of the marginal product of capital with respect to output is $-(1-\alpha)/\alpha$. If $\alpha = \frac{1}{3}$, a tenfold difference in output per worker arising from differences in capital per worker thus implies a hundredfold difference in the marginal product of capital. And since the return to capital is $f'(k) - \delta$, the difference in rates of return is even larger.

Again, there is no evidence of such differences in rates of return. Direct measurement of returns on financial assets, for example, suggests only moderate variation over time and across countries. More tellingly, we can learn much about cross-country differences simply by examining where the holders of capital want to invest. If rates of return were larger by a factor of 10 or 100 in poor countries than in rich countries, there would be immense incentives to invest in poor countries. Such differences in rates of return would swamp such considerations as capital-market imperfections, government tax policies, fear of expropriation, and so on, and we would observe immense flows of capital from rich to poor countries. We do not see such flows.[19]

Thus differences in physical capital per worker cannot account for the differences in output per worker that we observe, at least if capital's contribution to output is roughly reflected by its private returns.

The other potential source of variation in output per worker in the Solow model is the effectiveness of labor. Attributing differences in standards of living to differences in the effectiveness of labor does not require huge differences in capital or in rates of return. Along a balanced growth path, for example, capital is growing at the same rate as output; and the marginal product of capital, $f'(k)$, is constant.

The Solow model's treatment of the effectiveness of labor is highly incomplete, however. Most obviously, the growth of the effectiveness of labor is exogenous: the model takes as given the behavior of the variable that it identifies as the driving force of growth. Thus it is only a small exaggeration to say that we have been modeling growth by assuming it.

More fundamentally, the model does not identify what the "effectiveness of labor" is; it is just a catchall for factors other than labor and capital that affect output. Thus saying that differences in income are due to differences in the effectiveness of labor is no different than saying that they are not due to differences in capital per worker. To proceed, we must take a stand concerning what we mean by the effectiveness of labor and what causes it to vary. One natural possibility is that the effectiveness of labor corresponds to abstract knowledge. To understand worldwide growth, it would then be necessary to analyze the determinants of the stock of knowledge over time. To understand cross-country differences in real incomes, one would have to explain why firms in some countries have access to more knowledge than firms in other countries, and why that greater knowledge is not rapidly transmitted to poorer countries.

There are other possible interpretations of A: the education and skills of the labor force, the strength of property rights, the quality of infrastructure, cultural attitudes toward entrepreneurship and work, and so on. Or A may reflect a combination of forces. For any proposed view of what A represents, one would again have to address the questions of how it affects output, how it evolves over time, and why it differs across parts of the world.

The other possible way to proceed is to consider the possibility that capital is more important than the Solow model implies. If capital encompasses more than just physical capital, or if physical capital has positive

[19] One can try to avoid this conclusion by considering production functions where capital's marginal product falls less rapidly as k rises than it does in the Cobb-Douglas case. This approach encounters two major difficulties. First, since it implies that the marginal product of capital is similar in rich and poor countries, it implies that capital's share is much larger in rich countries. Second, and similarly, it implies that real wages are only slightly larger in rich than in poor countries. These implications appear grossly inconsistent with the facts.

externalities, then the private return on physical capital is not an accurate guide to capital's importance in production. In this case, the calculations we have done may be misleading, and it may be possible to resuscitate the view that differences in capital are central to differences in incomes.

These possibilities for addressing the fundamental questions of growth theory are the subject of Chapter 3.

1.7 Empirical Applications

Growth Accounting

In many situations, we are interested in the proximate determinants of growth. That is, we often want to know how much of growth over some period is due to increases in various factors of production, and how much stems from other forces. *Growth accounting,* which was pioneered by Abramovitz (1956) and Solow (1957), provides a way of tackling this subject.

To see how growth accounting works, consider again the production function $Y(t) = F(K(t), A(t)L(t))$. This implies

$$\dot{Y}(t) = \frac{\partial Y(t)}{\partial K(t)}\dot{K}(t) + \frac{\partial Y(t)}{\partial L(t)}\dot{L}(t) + \frac{\partial Y(t)}{\partial A(t)}\dot{A}(t). \tag{1.33}$$

$\partial Y/\partial L$ and $\partial Y/\partial A$ denote $[\partial Y/\partial(AL)]A$ and $[\partial Y/\partial(AL)]L$, respectively. Dividing both sides by $Y(t)$ and rewriting the terms on the right-hand side yields

$$\begin{aligned}\frac{\dot{Y}(t)}{Y(t)} &= \frac{K(t)}{Y(t)}\frac{\partial Y(t)}{\partial K(t)}\frac{\dot{K}(t)}{K(t)} + \frac{L(t)}{Y(t)}\frac{\partial Y(t)}{\partial L(t)}\frac{\dot{L}(t)}{L(t)} + \frac{A(t)}{Y(t)}\frac{\partial Y(t)}{\partial A(t)}\frac{\dot{A}(t)}{A(t)} \\ &\equiv \alpha_K(t)\frac{\dot{K}(t)}{K(t)} + \alpha_L(t)\frac{\dot{L}(t)}{L(t)} + R(t).\end{aligned} \tag{1.34}$$

Here $\alpha_L(t)$ is the elasticity of output with respect to labor at time t, $\alpha_K(t)$ is again the elasticity of output with respect to capital, and $R(t) \equiv [A(t)/Y(t)][\partial Y(t)/\partial A(t)][\dot{A}(t)/A(t)]$. Subtracting $\dot{L}(t)/L(t)$ from both sides and using the fact that $\alpha_L(t) + \alpha_K(t) = 1$ (see Problem 1.9) gives an expression for the growth rate of output per worker:

$$\frac{\dot{Y}(t)}{Y(t)} - \frac{\dot{L}(t)}{L(t)} = \alpha_K(t)\left[\frac{\dot{K}(t)}{K(t)} - \frac{\dot{L}(t)}{L(t)}\right] + R(t). \tag{1.35}$$

The growth rates of Y, K, and L are straightforward to measure. And we know that if capital earns its marginal product, α_K can be measured using data on the share of income that goes to capital. $R(t)$ can then be measured as the residual in (1.35). Thus (1.35) provides a way of decomposing the growth of output per worker into the contribution of growth of capital per worker and a remaining term, the *Solow residual.* The Solow residual is sometimes interpreted as a measure of the contribution of technological progress. As the derivation shows, however, it reflects all sources of growth other than the contribution of capital accumulation via its private return.

This basic framework can be extended in many ways. The most common extensions are to consider different types of capital and labor and to adjust for changes in the quality of inputs. But more complicated adjustments are also possible. For example, if there is evidence of imperfect competition, one can try to adjust the data on income shares to obtain a better estimate of the elasticity of output with respect to the different inputs.

Growth accounting only examines the immediate determinants of growth: it asks how much factor accumulation, improvements in the quality of inputs, and so on contribute to growth while ignoring the deeper issue of what causes the changes in those determinants. One way to see that growth accounting does not get at the underlying sources of growth is to consider what happens if it is applied to an economy described by the Solow model that is on its balanced growth path. We know that in this case growth is coming entirely from growth in A. But, as Problem 1.13 asks you to show and explain, growth accounting in this case attributes only fraction $1 - \alpha_K(k^*)$ of growth to the residual and fraction $\alpha_K(k^*)$ to capital accumulation.

Even though growth accounting provides evidence only about the immediate sources of growth, it has been fruitfully applied to many issues. For example, it has played a major role in a recent debate concerning the exceptionally rapid growth of the newly industrializing countries of East Asia. Young (1995) uses detailed growth accounting to argue that the higher growth in these countries than in the rest of the world is almost entirely due to rising investment, increasing labor force participation, and improving labor quality (in terms of education), and not to rapid technological progress and other forces affecting the Solow residual. This suggests that for other countries to replicate the NICs' successes, it is enough for them to promote accumulation of physical and human capital and greater use of resources, and that they need not tackle the even more difficult task of finding ways of obtaining greater output for a given set of inputs. In this view, the NICs' policies concerning trade, regulation, and so on have been important largely only to the extent they have influenced factor accumulation and factor use.

Hsieh (2002), however, observes that one can do growth accounting by examining the behavior of factor returns rather than quantities. If rapid growth comes solely from capital accumulation, for example, we will see either a large fall in the return to capital or a large rise in capital's share (or a combination). Doing the growth accounting this way, Hsieh finds a much larger role for the residual. But Young (1998) takes issue with Hsieh's analysis, and argues that the evidence from factor returns is in fact consistent with his original conclusions.

Growth accounting has also been used extensively to study both the productivity growth slowdown (the reduced growth rate of output per worker-hour in the United States and other industrialized countries that began in the early 1970s) and the productivity growth rebound (the return of U.S.

productivity growth starting in the mid-1990s to close to its level before the slowdown). Growth-accounting studies of the rebound suggest that computers and other types of information technology are the main source of the rebound (see, for example, Oliner and Sichel, 2002). Until the mid-1990s, the rapid technological progress in computers and their introduction in many sectors of the economy appear to have had little impact on aggregate productivity. In part, this was simply because computers, although spreading rapidly, were still only a small fraction of the overall capital stock. And in part, it was because the adoption of the new technologies involved substantial adjustment costs. Since the mid-1990s, however, computers and other forms of information technology have had a large impact on aggregate productivity.

At this point, computer use is still increasing rapidly, and computers represent a significant portion of the capital stock. As a result, even if technological progress in computers and information technology slows from its extraordinary rates of recent decades, further improvement and dissemination of information technology is likely to continue to contribute substantially to aggregate productivity growth for some time. Thus, as Oliner and Sichel describe, almost everyone who has studied the issue carefully believes that the most likely outcome is that the productivity growth rebound will be sustained in the United States for at least the next 5 or 10 years. Of course, productivity growth is very difficult to forecast; thus the actual outcome remains quite uncertain.[20]

Convergence

An issue that has attracted considerable attention in empirical work on growth is whether poor countries tend to grow faster than rich countries. There are at least three reasons that one might expect such convergence. First, the Solow model predicts countries converge to their balanced growth paths. Thus to the extent that differences in output per worker arise from countries being at different points relative to their balanced growth paths, one would expect poor countries to catch up to rich ones. Second, the Solow

[20] The simple information-technology explanation of the productivity growth rebound faces an important challenge, however: other industrialized countries have for the most part not shared in the rebound. The leading candidate explanation of this puzzle is closely related to the observation that there are large adjustments costs in adopting the new technologies. In this view, the adoption of computers and information technology raises productivity substantially only if it is accompanied by major changes in worker training, the composition of the firm's workforce, and the organization of the firm. Thus in countries where firms lack the ability to make these changes (because of either government regulation or business culture), the information-technology revolution is, as yet, having little impact on overall economic performance (see, for example, Breshnahan, Brynjolfsson, and Hitt, 2002, and Basu, Fernald, Oulton, and Srinivasan, 2003).

model implies that the rate of return on capital is lower in countries with more capital per worker. Thus there are incentives for capital to flow from rich to poor countries; this will also tend to cause convergence. And third, if there are lags in the diffusion of knowledge, income differences can arise because some countries are not yet employing the best available technologies. These differences might tend to shrink as poorer countries gain access to state-of-the-art methods.

Baumol (1986) examines convergence from 1870 to 1979 among the 16 industrialized countries for which Maddison (1982) provides data. Baumol regresses output growth over this period on a constant and initial income. That is, he estimates

$$\ln\left[\left(\frac{Y}{N}\right)_{i,1979}\right] - \ln\left[\left(\frac{Y}{N}\right)_{i,1870}\right] = a + b\ln\left[\left(\frac{Y}{N}\right)_{i,1870}\right] + \varepsilon_i. \tag{1.36}$$

Here $\ln(Y/N)$ is log income per person, ε is an error term, and i indexes countries.[21] If there is convergence, b will be negative: countries with higher initial incomes have lower growth. A value for b of -1 corresponds to perfect convergence: higher initial income on average lowers subsequent growth one-for-one, and so output per person in 1979 is uncorrelated with its value in 1870. A value for b of 0, on the other hand, implies that growth is uncorrelated with initial income and thus that there is no convergence.

The results are

$$\ln\left[\left(\frac{Y}{N}\right)_{i,1979}\right] - \ln\left[\left(\frac{Y}{N}\right)_{i,1870}\right] = 8.457 - \underset{(0.094)}{0.995}\ \ln\left[\left(\frac{Y}{N}\right)_{i,1870}\right], \tag{1.37}$$

$$R^2 = 0.87, \qquad \text{s.e.e.} = 0.15,$$

where the number in parentheses, 0.094, is the standard error of the regression coefficient. Figure 1.7 shows the scatterplot corresponding to this regression.

The regression suggests almost perfect convergence. The estimate of b is almost exactly equal to -1, and it is estimated fairly precisely; the two-standard-error confidence interval is (0.81, 1.18). In this sample, per capita income today is essentially unrelated to per capita income 100 years ago.

DeLong (1988) demonstrates, however, that Baumol's finding is largely spurious. There are two problems. The first is *sample selection.* Since historical data are constructed retrospectively, the countries that have long data series are generally those that are the most industrialized today. Thus countries that were not rich 100 years ago are typically in the sample only if they grew rapidly over the next 100 years. Countries that were rich 100 years

21 Baumol considers output per worker rather than output per person. This choice has little effect on the results.

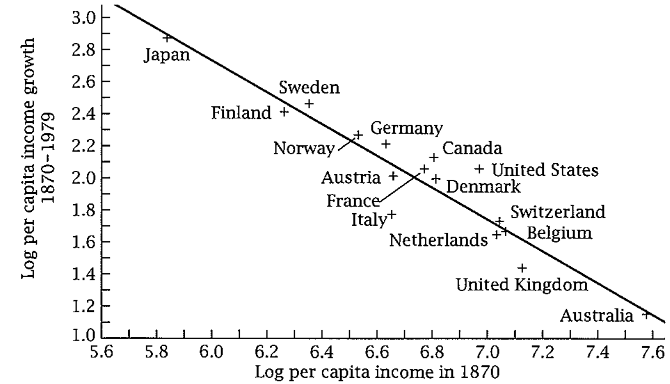

FIGURE 1.7 Initial income and subsequent growth in Baumol's sample (from DeLong, 1988; used with permission)

ago, in contrast, are generally included even if their subsequent growth was only moderate. Because of this, we are likely to see poorer countries growing faster than richer ones in the sample of countries we consider, even if there is no tendency for this to occur on average.

The natural way to eliminate this bias is to use a rule for choosing the sample that is not based on the variable we are trying to explain, which is growth over the period 1870–1979. Lack of data makes it impossible to include the entire world. DeLong therefore considers the richest countries as of 1870; specifically, his sample consists of all countries at least as rich as the second poorest country in Baumol's sample in 1870, Finland. This causes him to add seven countries to Baumol's list (Argentina, Chile, East Germany, Ireland, New Zealand, Portugal, and Spain) and to drop one (Japan).[22]

Figure 1.8 shows the scatterplot for the unbiased sample. The inclusion of the new countries weakens the case for convergence considerably. The regression now produces an estimate of b of -0.566, with a standard error of 0.144. Thus accounting for the selection bias in Baumol's procedure eliminates about half of the convergence that he finds.

The second problem that DeLong identifies is *measurement error.* Estimates of real income per capita in 1870 are imprecise. Measurement

22 Since a large fraction of the world was richer than Japan in 1870, it is not possible to consider all countries at least as rich as Japan. In addition, one has to deal with the fact that countries' borders are not fixed. DeLong chooses to use 1979 borders. Thus his 1870 income estimates are estimates of average incomes in 1870 in the geographic regions defined by 1979 borders.

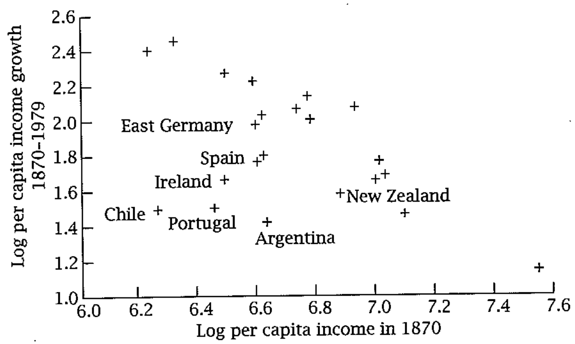

FIGURE 1.8 Initial income and subsequent growth in the expanded sample (from DeLong, 1988; used with permission)

error again creates bias toward finding convergence. When 1870 income is overstated, growth over the period 1870–1979 is understated by an equal amount; when 1870 income is understated, the reverse occurs. Thus measured growth tends to be lower in countries with higher measured initial income even if there is no relation between actual growth and actual initial income.

DeLong therefore considers the following model:

$$\ln\left[\left(\frac{Y}{N}\right)_{i,1979}\right] - \ln\left[\left(\frac{Y}{N}\right)_{i,1870}\right]^* = a + b\ln\left[\left(\frac{Y}{N}\right)_{i,1870}\right]^* + \varepsilon_i, \tag{1.38}$$

$$\ln\left[\left(\frac{Y}{N}\right)_{i,1870}\right] = \ln\left[\left(\frac{Y}{N}\right)_{i,1870}\right]^* + u_i. \tag{1.39}$$

Here $\ln[(Y/N)_{1870}]^*$ is the true value of log income per capita in 1870 and $\ln[(Y/N)_{1870}]$ is the measured value. ε and u are assumed to be uncorrelated with each other and with $\ln[(Y/N)_{1870}]^*$.

Unfortunately, it is not possible to estimate this model using only data on $\ln[(Y/N)_{1870}]$ and $\ln[(Y/N)_{1979}]$. The problem is that there are different hypotheses that make identical predictions about the data. For example, suppose we find that measured growth is negatively related to measured initial income. This is exactly what one would expect either if measurement error is unimportant and there is true convergence or if measurement error is important and there is no true convergence. Technically, the model is *not identified.*

DeLong argues, however, that we have at least a rough idea of how good the 1870 data are, and thus have a sense of what is a reasonable value

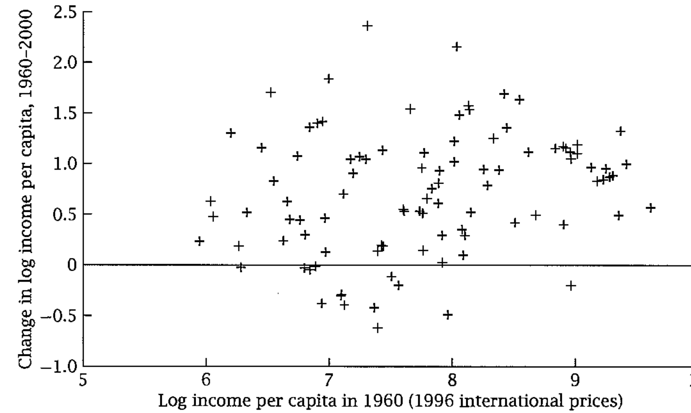

FIGURE 1.9 Initial income and subsequent growth in the postwar period

for the standard deviation of the measurement error. For example, $\sigma_u = 0.01$ implies that we have measured initial income to within an average of 1 percent; this is implausibly low. Similarly, $\sigma_u = 0.50$—an average error of 50 percent—seems implausibly high. DeLong shows that if we fix a value of σ_u, we can estimate the remaining parameters.

Even moderate measurement error has a substantial impact on the results. For the unbiased sample, the estimate of b reaches 0 (no tendency toward convergence) for $\sigma_u \simeq 0.15$, and is 1 (tremendous divergence) for $\sigma_u \simeq 0.20$. Thus plausible amounts of measurement error eliminate most or all of the remainder of Baumol's estimate of convergence.

It is also possible to investigate convergence for different samples of countries and different time periods. Figure 1.9 is a *convergence scatterplot* analogous to Figures 1.7 and 1.8 for virtually the entire non-Communist world for the period 1960–2000. As the figure shows, there is little evidence of convergence. We return to the issue of convergence in Section 3.12.

Saving and Investment

Consider a world where every country is described by the Solow model and where all countries have the same amount of capital per unit of effective labor. Now suppose that the saving rate in one country rises. If all the additional saving is invested domestically, the marginal product of capital in that country falls below that in other countries. The country's residents therefore have incentives to invest abroad. Thus if there are no impediments to capital flows, not all the additional saving is invested domestically. Instead, the investment resulting from the increased saving is spread uniformly over the whole world; the fact that the rise in saving occurred in one country has no special effect on investment there. Thus in the absence of barriers to capital movements, there is no reason to expect countries with high saving to also have high investment.

Feldstein and Horioka (1980) examine the association between saving and investment rates. They find that, contrary to this simple view, saving and investment rates are strongly correlated. Specifically, Feldstein and Horioka run a cross-country regression for 21 industrialized countries of the average share of investment in GDP during the period 1960–1974 on a constant and the average share of saving in GDP over the same period. The results are

$$\left(\frac{I}{Y}\right)_i = \underset{(0.018)}{0.035} + \underset{(0.074)}{0.887}\left(\frac{S}{Y}\right)_i, \qquad R^2 = 0.91, \tag{1.40}$$

where again the numbers in parentheses are standard errors. Thus, rather than there being no relation between saving and investment, there is an almost one-to-one relation.

There are various possible explanations for Feldstein and Horioka's finding. One possibility, suggested by Feldstein and Horioka, is that there are significant barriers to capital mobility. In this case, differences in saving and investment across countries would be associated with rate-of-return differences. There is little evidence of such rate-of-return differences, however.

Another possibility is that there are underlying variables that affect both saving and investment. For example, high tax rates can reduce both saving and investment (Barro, Mankiw, and Sala-i-Martin, 1995). Similarly, countries whose citizens have low discount rates, and thus high saving rates, may provide favorable investment climates in ways other than the high saving; for example, they may limit workers' ability to form strong unions.

Finally, the strong association between saving and investment can arise from government policies that offset forces that would otherwise make saving and investment differ. Governments may be averse to large gaps between saving and investment—after all, a large gap must be associated with a large trade deficit (if investment exceeds saving) or a large trade surplus (if saving exceeds investment). If economic forces would otherwise give rise to a large imbalance between saving and investment, the government may choose to adjust its own saving behavior or its tax treatment of saving or investment to bring them into rough balance. Helliwell (1998) finds that the saving-investment correlation is much weaker if we look across regions within a country rather than across countries. This is certainly consistent with the hypothesis that national governments take steps to prevent large imbalances between aggregate saving and investment, but that such imbalances can develop in the absence of government intervention.

In sum, the strong relationship between saving and investment differs dramatically from the predictions of a natural baseline model. Most likely,

however, this difference reflects not major departures from the baseline (such as large barriers to capital mobility), but something less fundamental (such as underlying forces affecting both saving and investment).

1.8 The Environment and Economic Growth

Natural resources, pollution, and other environmental considerations are absent from the Solow model. But at least since Malthus (1798) made his classic argument, many people have believed that these considerations are critical to the possibilities for long-run economic growth. For example, the amounts of oil and other natural resources on earth are fixed. This could mean that any attempt to embark on a path of perpetually rising output will eventually deplete those resources, and must therefore fail. Similarly, the fixed supply of land may become a binding constraint on our ability to produce. Or ever-increasing output may generate an ever-increasing stock of pollution that will bring growth to a halt.[23]

This section addresses the issue of how environmental limitations affect long-run growth. In thinking about this issue, it is important to distinguish between environmental factors for which there are well-defined property rights—notably natural resources and land—and those for which there are not—notably pollution-free air and water.

The existence of property rights for an environmental good has two important implications. The first is that markets provide valuable signals concerning how the good should be used. Suppose, for example, that the best available evidence indicates that the limited supply of oil will be an important limitation on our ability to produce in the future. This means that oil will command a high price in the future. But this in turn implies that the owners of oil do not want to sell their oil cheaply today. Thus oil commands a high price today, and so current users have an incentive to conserve. In short, evidence that the fixed amount of oil is likely to limit our ability to produce in the future would not be grounds for government intervention. Such a situation, though unfortunate, would be addressed by the market.

The second implication of the existence of property rights for an environmental good is that we can use the good's price to obtain evidence about its importance in production. For example, since evidence that oil will be an important constraint on future production would cause it to have a high price today, economists can use the current price to infer what the best available evidence suggests about oil's importance; they do not need to assess that evidence independently.

23 An influential modern statement of these concerns is Meadows, Meadows, Randers, and Behrens (1972).

With environmental goods for which there are no property rights, the use of a good has externalities. For example, firms can pollute without compensating the people they harm. Thus the case for government intervention is much stronger. And there is no market price to provide a handy summary of the evidence concerning the good's importance. As a result, economists interested in environmental issues must attempt to assess that evidence themselves.

We will begin by considering environmental goods that are traded in markets. We will analyze both a simple baseline case and an important complication to the baseline. We will then turn to environmental goods for which there is no well-functioning market.

Natural Resources and Land: A Baseline Case

We want to extend our analysis to include natural resources and land. To keep the analysis manageable, we start with the case of Cobb-Douglas production. Thus the production function, (1.1), becomes

$$Y(t) = K(t)^{\alpha} R(t)^{\beta} T(t)^{\gamma} [A(t)L(t)]^{1-\alpha-\beta-\gamma},$$
$$\alpha > 0, \quad \beta > 0, \quad \gamma > 0, \quad \alpha+\beta+\gamma < 1. \tag{1.41}$$

Here R denotes resources used in production, and T denotes the amount of land.

The dynamics of capital, labor, and the effectiveness of labor are the same as before: $\dot{K}(t) = sY(t) - \delta K(t)$, $\dot{L}(t) = nL(t)$, and $\dot{A}(t) = gA(t)$. The new assumptions concern resources and land. Since the amount of land on earth is fixed, in the long run the quantity used in production cannot be growing. Thus we assume

$$\dot{T}(t) = 0. \tag{1.42}$$

Similarly, the facts that resource endowments are fixed and that resources are used in production imply that resource use must eventually decline. Thus, even though resource use has been rising historically, we assume

$$\dot{R}(t) = -bR(t), \qquad b > 0. \tag{1.43}$$

The presence of resources and land in the production function means that K/AL no longer converges to some value. As a result, we cannot use our previous approach of focusing on K/AL to analyze the behavior of this economy. A useful strategy in such situations is to ask whether there can be a balanced growth path and, if so, what the growth rates of the economy's variables are on that path.

By assumption, A, L, R, and T are each growing at a constant rate. Thus what is needed for a balanced growth path is that K and Y each grow at a constant rate. The equation of motion for capital, $\dot{K}(t) = sY(t) - \delta K(t)$, implies that the growth rate of K is

$$\frac{\dot{K}(t)}{K(t)} = s\frac{Y(t)}{K(t)} - \delta. \tag{1.44}$$

Thus for the growth rate of K to be constant, Y/K must be constant. That is, the growth rates of Y and K must be equal.

We can use the production function, (1.41), to find when this can occur. Taking logs of both sides of (1.41) gives us

$$\ln Y(t) = \alpha \ln K(t) + \beta \ln R(t) + \gamma \ln T(t) + (1-\alpha-\beta-\gamma)[\ln A(t) + \ln L(t)]. \tag{1.45}$$

We can now differentiate both sides of this expression with respect to time. Using the fact that the time derivative of the log of a variable equals the variable's growth rate, we obtain

$$g_Y(t) = \alpha g_K(t) + \beta g_R(t) + \gamma g_T(t) + (1-\alpha-\beta-\gamma)[g_A(t) + g_L(t)], \tag{1.46}$$

where g_X denotes the growth rate of X. The growth rates of R, T, A, and L are $-b$, 0, g, and n, respectively. Thus (1.46) simplifies to

$$g_Y(t) = \alpha g_K(t) - \beta b + (1-\alpha-\beta-\gamma)(n+g). \tag{1.47}$$

We can now use our finding that g_Y and g_K must be equal if the economy is on a balanced growth path. Imposing $g_K = g_Y$ on (1.47) and solving for g_Y gives us

$$g_Y^{bgp} = \frac{(1-\alpha-\beta-\gamma)(n+g) - \beta b}{1-\alpha}, \tag{1.48}$$

where g_Y^{bgp} denotes the growth rate of Y on the balanced growth path.

This analysis leaves out a step: we have not determined whether the economy in fact converges to this balanced growth path. From (1.47), we know that if g_K exceeds its balanced-growth-path value, g_Y does as well, but by less than g_K does. Thus if g_K exceeds its balanced-growth-path value, Y/K is falling. Equation (1.44) tells us that g_K equals $s(Y/K) - \delta$. Thus if Y/K is falling, g_K is falling as well. That is, if g_K exceeds its balanced-growth-path value, it is falling. Similarly, if it is less than its balanced-growth-path value, it is rising. Thus g_K converges to its balanced-growth-path value, and so the economy converges to its balanced growth path.[24]

[24] This analysis overlooks one subtlety. If $(1-\alpha-\beta-\gamma)(n+g) + (1-\alpha)\delta - \beta b$ is negative, the condition $g_K = g_K^{bgp}$ holds only for a negative value of Y/K. And the statement that Y/K is falling when g_Y is less than g_K is not true if Y/K is zero or negative. As a result, if $(1-\alpha-\beta-\gamma)(n+g) + (1-\alpha)\delta - \beta b$ is negative, the economy does not converge to the balanced growth path described in the text, but to a situation where $Y/K = 0$ and $g_K = -\delta$. But for any reasonable parameter values, $(1-\alpha-\beta-\gamma)(n+g) + (1-\alpha)\delta - \beta b$ is positive. Thus this complication is not important.

Equation (1.48) implies that the growth rate of output per worker on the balanced growth path is

$$
\begin{aligned}
g_{Y/L}^{bgp} &= g_Y^{bgp} - g_L^{bgp} \\
&= \frac{(1-\alpha-\beta-\gamma)(n+g)-\beta b}{1-\alpha} - n \qquad (1.49) \\
&= \frac{(1-\alpha-\beta-\gamma)g-\beta b-(\beta+\gamma)n}{1-\alpha}.
\end{aligned}
$$

Equation (1.49) shows that growth in income per worker on the balanced growth path, $g_{Y/L}^{bgp}$, can be either positive or negative. That is, resource and land limitations can cause output per worker to eventually be falling, but they need not. The declining quantities of resources and land per worker are drags on growth. But technological progress is a spur to growth. If the spur created by technological progress is larger than the drags exerted by resources and land, then there is sustained growth in output per worker. This is precisely what has happened over the past few centuries.

An Illustrative Calculation

In recent history, the advantages of technological progress have outweighed the disadvantages of resource and land limitations. But this does not tell us how large those disadvantages are. For example, they might be large enough that only a moderate slowing of technological progress would make overall growth in income per worker negative.

Resource and land limitations reduce growth by causing resource use per worker and land per worker to be falling. Thus, as Nordhaus (1992) observes, to gauge how much these limitations are reducing growth, we need to ask how much greater growth would be if resources and land per worker were constant. Concretely, consider an economy identical to the one we have just considered except that the assumptions $\dot{T}(t) = 0$ and $\dot{R}(t) = -bR(t)$ are replaced with the assumptions $\dot{T}(t) = nT(t)$ and $\dot{R}(t) = nR(t)$. In this hypothetical economy, there are no resource and land limitations; both grow as population grows. Analysis parallel to that used to derive equation (1.49) shows that growth of output per worker on the balanced growth path of this economy is[25]

$$\tilde{g}_{Y/L}^{bgp} = \frac{1}{1-\alpha}(1-\alpha-\beta-\gamma)g. \qquad (1.50)$$

[25] See Problem 1.15.

The "growth drag" from resource and land limitations is the difference between growth in this hypothetical case and growth in the case of resource and land limitations:

$$
\begin{aligned}
\text{Drag} &= \tilde{g}_{Y/L}^{bgp} - g_{Y/L}^{bgp} \\
&= \frac{(1-\alpha-\beta-\gamma)g-[(1-\alpha-\beta-\gamma)g-\beta b-(\beta+\gamma)n]}{1-\alpha} \qquad (1.51) \\
&= \frac{\beta b+(\beta+\gamma)n}{1-\alpha}.
\end{aligned}
$$

Thus, the growth drag is increasing in resources' share (β), land's share (γ), the rate that resource use is falling (b), the rate of population growth (n), and capital's share (α).

It is possible to quantify the size of the drag. Because resources and land are traded in markets, we can use income data to estimate their importance in production—that is, to estimate β and γ. As Nordhaus (1992) describes, these data suggest a combined value of $\beta+\gamma$ of about 0.2. Nordhaus goes on to use a somewhat more complicated version of the framework presented here to estimate the growth drag. His point estimate is a drag of 0.0024—that is, about a quarter of a percentage point per year. He finds that only about a quarter of the drag is due to the limited supply of land. Of the remainder, he estimates that the vast majority is due to limited energy resources.

Thus this evidence suggests that the reduction in growth caused by environmental limitations, while not trivial, is not large. In addition, since growth in income per worker has been far more than a quarter of a percentage point per year, the evidence suggests that there would have to be very large changes for resource and land limitations to cause income per worker to start falling.

A Complication

The stock of land is fixed, and resource use must eventually fall. Thus even though technology has been able to keep ahead of resource and land limitations over the past few centuries, it may still appear that those limitations must eventually become a binding constraint on our ability to produce.

The reason that this does not occur in our model is that production is Cobb-Douglas. With Cobb-Douglas production, a given percentage change in A always produces the same percentage change in output, regardless of how large A is relative to R and T. As a result, technological progress can always counterbalance declines in R/L and T/L.

This is not a general property of production functions, however. With Cobb–Douglas production, the elasticity of substitution between inputs is 1. If this elasticity is less than 1, the share of income going to the inputs that are becoming scarcer rises over time. Intuitively, as the production function becomes more like the Leontief case, the inputs that are becoming scarcer become increasingly important. Conversely, if the elasticity of substitution is greater than 1, the share of income going to the inputs that are becoming scarcer is falling. This, too, is intuitive: as the production function becomes closer to linear, the abundant factors benefit.

In terms of our earlier analysis, what this means is that if we do not restrict our attention to Cobb–Douglas production, the shares in expression (1.51) for the growth drag are no longer constant, but are functions of factor proportions. And if the elasticity of substitution is less than 1, the share of income going to resources and land is rising over time—and thus the growth drag is as well. Indeed, in this case the share of income going to the slowest-growing input—resources—approaches 1. Thus the growth drag approaches $b+n$. That is, asymptotically income per worker declines at rate $b+n$, the rate at which resource use per worker is falling. This case supports our apocalyptic intuition: in the long run, the fixed supply of resources leads to steadily declining incomes.

In fact, however, recognizing that production may not be Cobb–Douglas should not raise our estimate of the importance of resource and land limitations, but reduce it. The reason is that the shares of income going to resources and land are falling rather than rising. We can write land's share as the real rental price of land multiplied by the ratio of land to output. The real rental price shows little trend, while the land-to-GDP ratio has been falling steadily. Thus land's share has been declining. Similarly, real resource prices have been falling moderately, and the ratio of resource use to GDP has also been falling. Thus resources' share has also been declining. And declining resource and land shares imply a falling growth drag.

The fact that land's and resources' shares have been declining despite the fact that these factors have been becoming relatively scarcer means that the elasticity of substitution between these inputs and the others must be greater than 1. At first glance, this may seem surprising. If we think in terms of narrowly defined goods—books, for example—possibilities for substitution among inputs may not seem particularly large. But if we recognize that what people value is not particular goods but the ultimate services they provide—information storage, for example—the idea that there are often large possibilities for substitution becomes more plausible. Information can be stored not only through books, but through oral tradition, stone tablets, microfilm, videotape, and disks. These different means of storage use capital, resources, land, and labor in very different proportions. As a result, the economy can respond to the increasing scarcity of resources and land by moving to means of information storage that use resources and land less intensively.

Pollution

Declining quantities of resources and land per worker are not the only ways that environmental problems can limit growth. Production creates pollution. This pollution reduces properly measured output. That is, if our data on real output accounted for all the outputs of production at prices that reflect their impacts on utility, pollution would enter with a negative price. In addition, pollution could rise to the point where it reduces conventionally measured output. For example, global warming could reduce output through its impact on sea levels and weather patterns.

Economic theory does not give us reason to be sanguine about pollution. Because those who pollute do not bear the costs of their pollution, an unregulated market leads to excessive pollution. Similarly, there is nothing to prevent an environmental catastrophe in an unregulated market. For example, suppose there is some critical level of pollution that would result in a sudden and drastic change in climate. Because pollution's effects are external, there is no market mechanism to prevent pollution from rising to such a level, or even a market price of a pollution-free environment to warn us that well-informed individuals believe a catastrophe is imminent.

Conceptually, the correct policy to deal with pollution is straightforward. We should estimate the dollar value of the negative externality and tax pollution by this amount. This would bring private and social costs in line, and thus would result in the socially optimal level of pollution.[26]

Although describing the optimal policy is easy, it is still useful to know how severe the problems posed by pollution are. In terms of understanding economic growth, we would like to know by how much pollution is likely to retard growth if no corrective measures are taken. In terms of policy, we would like to know how large a pollution tax is appropriate. We would also like to know whether, if pollution taxes are politically infeasible, the benefits of cruder regulatory approaches are likely to outweigh their costs. Finally, in terms of our own behavior, we would like to know how much effort individuals who care about others' well-being should make to curtail their activities that cause pollution.

Since there are no market prices to use as guides, economists interested in pollution must begin by looking at the scientific evidence. In the case of global warming, for example, a reasonable point estimate is that in the absence of major intervention, the average temperature will rise by 3 degrees centigrade over the period 1990–2050, with various effects on climate (Nordhaus, 1992). Economists can help estimate the welfare consequences of these changes. To give just one example, experts on farming had estimated the likely impact of global warming on U.S. farmers'

[26] Alternatively, we could find the socially optimal level of pollution and auction off a quantity of tradable permits that allow that level of pollution. Weitzman (1974) provides the classic analysis of the choice between controlling prices or quantities.

ability to continue growing their current crops. These studies concluded that global warming would have a significant negative impact. Mendelsohn, Nordhaus, and Shaw (1994), however, note that farmers can respond to changing weather patterns by moving into different crops, or even switching their land use out of crops altogether. They find that once these possibilities for substitution are taken into account, the overall effect of global warming on U.S. farmers is small and may be positive (see also Deschenes and Greenstone, 2004).

After considering the various channels through which global warming is likely to affect welfare, Nordhaus (1991) concludes that a reasonable estimate is that the overall welfare effect as of 2050 is likely to be slightly negative—the equivalent of a reduction in GDP of 1 to 2 percent. This corresponds to a reduction in average annual growth over the period 1990–2050 of only about 0.03 percentage points. Not surprisingly, Nordhaus finds that drastic measures to combat global warming, such as policies that would largely halt further warming by cutting emissions of greenhouse gases by 50 percent or more, would be much more harmful than simply doing nothing.

Using a similar approach, Nordhaus (1992) concludes that the welfare costs of other types of pollution are larger, but still limited. His point estimate is that they will lower appropriately measured annual growth by roughly 0.04 percentage points.

Of course, it is possible that this reading of the scientific evidence or this effort to estimate welfare effects is far from the mark. It is also possible that considering horizons longer than the 50 to 100 years usually examined in such studies would change the conclusions substantially. But the fact remains that most economists who have studied environmental issues seriously, even ones whose initial positions were sympathetic to environmental concerns, have concluded that the likely impact of environmental problems on growth is at most moderate.[27]

Problems

1.1. Basic properties of growth rates. Use the fact that the growth rate of a variable equals the time derivative of its log to show:

(*a*) The growth rate of the product of two variables equals the sum of their growth rates. That is, if $Z(t) = X(t)Y(t)$, then $\dot{Z}(t)/Z(t) = [\dot{X}(t)/X(t)] + [\dot{Y}(t)/Y(t)]$.

(*b*) The growth rate of the ratio of two variables equals the difference of their growth rates. That is, if $Z(t) = X(t)/Y(t)$, then $\dot{Z}(t)/Z(t) = [\dot{X}(t)/X(t)] - [\dot{Y}(t)/Y(t)]$.

(*c*) If $Z(t) = X(t)^\alpha$, then $\dot{Z}(t)/Z(t) = \alpha\dot{X}(t)/X(t)$.

1.2. Suppose that the growth rate of some variable, X, is constant and equal to $a > 0$ from time 0 to time t_1; drops to 0 at time t_1; rises gradually from 0 to a from time t_1 to time t_2; and is constant and equal to a after time t_2.

(*a*) Sketch a graph of the growth rate of X as a function of time.

(*b*) Sketch a graph of $\ln X$ as a function of time.

1.3. Describe how, if at all, each of the following developments affects the break-even and actual investment lines in our basic diagram for the Solow model:

(*a*) The rate of depreciation falls.

(*b*) The rate of technological progress rises.

(*c*) The production function is Cobb–Douglas, $f(k) = k^\alpha$, and capital's share, α, rises.

(*d*) Workers exert more effort, so that output per unit of effective labor for a given value of capital per unit of effective labor is higher than before.

1.4. Consider an economy with technological progress but without population growth that is on its balanced growth path. Now suppose there is a one-time jump in the number of workers.

(*a*) At the time of the jump, does output per unit of effective labor rise, fall, or stay the same? Why?

(*b*) After the initial change (if any) in output per unit of effective labor when the new workers appear, is there any further change in output per unit of effective labor? If so, does it rise or fall? Why?

(*c*) Once the economy has again reached a balanced growth path, is output per unit of effective labor higher, lower, or the same as it was before the new workers appeared? Why?

1.5. Suppose that the production function is Cobb–Douglas.

(*a*) Find expressions for k^*, y^*, and c^* as functions of the parameters of the model, s, n, δ, g, and α.

(*b*) What is the golden-rule value of k?

(*c*) What saving rate is needed to yield the golden-rule capital stock?

1.6. Consider a Solow economy that is on its balanced growth path. Assume for simplicity that there is no technological progress. Now suppose that the rate of population growth falls.

(*a*) What happens to the balanced-growth-path values of capital per worker, output per worker, and consumption per worker? Sketch the paths of these variables as the economy moves to its new balanced growth path.

(*b*) Describe the effect of the fall in population growth on the path of output (that is, total output, not output per worker).

[27] This does not imply that environmental factors are always unimportant to long-run growth. Brander and Taylor (1998) make a strong case that Easter Island suffered an environmental disaster of the type envisioned by Malthusians sometime between its settlement around 400 and the arrival of Europeans in the 1700s. And they argue that other primitive societies may have also suffered such disasters.

1.7. Find the elasticity of output per unit of effective labor on the balanced growth path, y^*, with respect to the rate of population growth, n. If $\alpha_K(k^*) = \frac{1}{3}$, $g = 2\%$, and $\delta = 3\%$, by about how much does a fall in n from 2 percent to 1 percent raise y^*?

1.8. Suppose that investment as a fraction of output in the United States rises permanently from 0.15 to 0.18. Assume that capital's share is $\frac{1}{3}$.

(*a*) By about how much does output eventually rise relative to what it would have been without the rise in investment?

(*b*) By about how much does consumption rise relative to what it would have been without the rise in investment?

(*c*) What is the immediate effect of the rise in investment on consumption? About how long does it take for consumption to return to what it would have been without the rise in investment?

1.9. Factor payments in the Solow model. Assume that both labor and capital are paid their marginal products. Let w denote $\partial F(K,AL)/\partial L$ and r denote $[\partial F(K,AL)/\partial K] - \delta$.

(*a*) Show that the marginal product of labor, w, is $A[f(k) - kf'(k)]$.

(*b*) Show that if both capital and labor are paid their marginal products, constant returns to scale imply that the total amount paid to the factors of production equals total net output. That is, show that under constant returns, $wL + rK = F(K,AL) - \delta K$.

(*c*) The return to capital (r) is roughly constant over time, as are the shares of output going to capital and to labor. Does a Solow economy on a balanced growth path exhibit these properties? What are the growth rates of w and r on a balanced growth path?

(*d*) Suppose the economy begins with a level of k less than k^*. As k moves toward k^*, is w growing at a rate greater than, less than, or equal to its growth rate on the balanced growth path? What about r?

1.10. Suppose that, as in Problem 1.9, capital and labor are paid their marginal products. In addition, suppose that all capital income is saved and all labor income is consumed. Thus $\dot{K} = [\partial F(K,AL)/\partial K]K - \delta K$.

(*a*) Show that this economy converges to a balanced growth path.

(*b*) Is k on the balanced growth path greater than, less than, or equal to the golden-rule level of k? What is the intuition for this result?

1.11. Go through steps analogous to those in equations (1.28)–(1.31) to find how quickly y converges to y^* in the vicinity of the balanced growth path. (Hint: Since $y = f(k)$, we can write $k = g(y)$, where $g(\bullet) = f^{-1}(\bullet)$.)

1.12. Embodied technological progress. (This follows Solow, 1960, and Sato, 1966.) One view of technological progress is that the productivity of capital goods built at t depends on the state of technology at t and is unaffected by subsequent technological progress. This is known as *embodied technological progress* (technological progress must be "embodied" in new capital before it can raise output). This problem asks you to investigate its effects.

(*a*) As a preliminary, let us modify the basic Solow model to make technological progress capital-augmenting rather than labor-augmenting. So that a balanced growth path exists, assume that the production function is Cobb–Douglas: $Y(t) = [A(t)K(t)]^\alpha L(t)^{1-\alpha}$. Assume that A grows at rate μ: $\dot{A}(t) = \mu A(t)$.

Show that the economy converges to a balanced growth path, and find the growth rates of Y and K on the balanced growth path. (Hint: Show that we can write $Y/(A^\phi L)$ as a function of $K/(A^\phi L)$, where $\phi = \alpha/(1-\alpha)$. Then analyze the dynamics of $K/(A^\phi L)$.)

(*b*) Now consider embodied technological progress. Specifically, let the production function be $Y(t) = J(t)^\alpha L(t)^{1-\alpha}$, where $J(t)$ is the effective capital stock. The dynamics of $J(t)$ are given by $\dot{J}(t) = sA(t)Y(t) - \delta J(t)$. The presence of the $A(t)$ term in this expression means that the productivity of investment at t depends on the technology at t.

Show that the economy converges to a balanced growth path. What are the growth rates of Y and J on the balanced growth path? (Hint: Let $\tilde{J}(t) = J(t)/A(t)$. Then use the same approach as in (*a*), focusing on $\tilde{J}/(A^\phi L)$ instead of $K/(A^\phi L)$.)

(*c*) What is the elasticity of output on the balanced growth path with respect to s?

(*d*) In the vicinity of the balanced growth path, how rapidly does the economy converge to the balanced growth path?

(*e*) Compare your results for (*c*) and (*d*) with the corresponding results in the text for the basic Solow model.

1.13. Consider a Solow economy on its balanced growth path. Suppose the growth-accounting techniques described in Section 1.7 are applied to this economy.

(*a*) What fraction of growth in output per worker does growth accounting attribute to growth in capital per worker? What fraction does it attribute to technological progress?

(*b*) How can you reconcile your results in (*a*) with the fact that the Solow model implies that the growth rate of output per worker on the balanced growth path is determined solely by the rate of technological progress?

1.14. (*a*) In the model of convergence and measurement error in equations (1.38) and (1.39), suppose the true value of b is -1. Does a regression of $\ln(Y/N)_{1979} - \ln(Y/N)_{1870}$ on a constant and $\ln(Y/N)_{1870}$ yield a biased estimate of b? Explain.

(*b*) Suppose there is measurement error in measured 1979 income per capita but not in 1870 income per capita. Does a regression of $\ln(Y/N)_{1979} - \ln(Y/N)_{1870}$ on a constant and $\ln(Y/N)_{1870}$ yield a biased estimate of b? Explain.

1.15. Derive equation (1.50). (Hint: Follow steps analogous to those in equations [1.47] and [1.48].)