4 Theories of Economic Growth

Economists have long puzzled over the question of economic growth. What is it that makes some countries rich while others remain poor? Formal studies of this question date back at least to the eighteenth-century writings of Scottish social philosopher Adam Smith. The search for answers continues to dominate economic thinking. In a lecture presented in 1985, Nobel Prize-winning economist Robert Lucas noted the then-rapid economic growth of Indonesia and Egypt and slow growth of India, famously asking,

> Is there some action a government of India could take that would lead the Indian economy to grow like Indonesia's or Egypt's? If so,what exactly? If not, what is it about the "nature of India" that makes it so? The consequences for human welfare involved in questions like these are simply staggering: Once one starts to think about them it is hard to think of anything else.[1]

More than a quarter century later, economists Dani Rodrik, Arvind Subramanian, and Francesco Trebbi noted,

> Average income levels in the world's richest and poorest nations differ by a factor of more than 100. Sierra Leone, the poorest economy for which we have national income statistics, has a per-capita GDP of $490, compared to Luxembourg's $50,061. What accounts for these differences, and what (if anything) can we do

[1]Robert E. Lucas, "On the Mechanics of Economic Development," *Journal of Monetary Economics* 22, no. 1 (July 1988), 3–42.

to reduce them? It is hard to think of any question in economics that is of greater intellectual significance, or of greater relevance to the vast majority of the world's population.[2]

Two observations emerge from the juxtaposition of these strikingly similar quotations: (1) Long-run economic growth may be the single most fundamental determinant of human welfare around the world, and (2) despite substantial efforts and significant progress toward solving the puzzle of economic growth during the 27 years that separate these comments, we remain far from a complete and policy-relevant understanding of the deep determinants of growth.

This chapter explores key contributions to the theory of economic growth. We began to explore these issues in the last chapter by examining some of the basic processes and patterns that characterize economic growth in low-income countries. We emphasized that growth depends on two processes: the *accumulation of assets* (such as capital, labor, and land), and *making those assets more productive.* Saving and investment are central, but investments must be productive for growth to proceed. Our approach was largely empirical, as we examined much of the data on growth and some of the key findings from research on the determinants of growth across countries. We saw that government policy, institutions, political and economic stability, geography, natural resource endowments, and levels of health and education all play some role in influencing economic growth. We emphasized that growth is not the same as development, but it remains absolutely central to the development process.

This chapter develops these ideas more formally by introducing the underlying theory and the most important basic models of economic growth that influence development thinking today. These models provide consistent frameworks for understanding the growth process and provide a theoretical foundation for the empirical approach we took in the last chapter. Here, we identify specific mathematical relationships between the quantity of capital and labor, their productivity, and the resulting aggregate output. It is important that these models also explore the process of accumulating *additional* capital and labor and *increasing* their productivity, which shifts the model from determining the *level* of output to the *rate of change* of output, which of course is the rate of economic growth.

As we begin to examine the models, it is useful to consider the words of Robert Solow, the father of modern growth theory, who once wrote: "All theory depends on assumptions that are not quite true. That is what makes it theory. The art of successful theorizing is to make the inevitable simplifying assumptions in such a way that

[2]Dani Rodrik, Arvind Subramanian, and Francesco Trebbi, "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development," *Journal of Economic Growth*, 9, no. 2 (2004), 131-65.

the final results are not very sensitive."[3] The best models are simple, yet still manage to communicate powerful insights into how the real world operates. In this spirit, the models presented here make assumptions that clearly are not true but allow us to simplify the framework and make it easier to grasp key concepts and insights. For example, we begin by assuming that our prototype economy has one type of homogeneous worker and one type of capital good that combine to produce one standard product. No economy in the world has characteristics even closely resembling these assumptions, but making these assumptions allows us to cut through many details and get to the core concepts of the theory of economic growth.

THE BASIC GROWTH MODEL

The most fundamental models of economic output and economic growth are based on a small number of equations that relate saving, investment, and population growth to the size of the workforce and capital stock and, in turn, to aggregate production of a single good. These models initially focus on the *levels* of investment, labor, productivity, and output. It then becomes straightforward to examine the *changes* in these variables. Our ultimate focus is to explore the key determinants of the *change in output*—that is, on the rate of economic growth. The version of the basic model that we examine here has five equations: an aggregate production function, an equation determining the level of saving, the saving-investment identity, a statement relating new investment to changes in the capital stock, and an expression for the growth rate of the labor force.[4] We examine each of these in turn.

Standard growth models have at their core a production function. At the individual firm or microeconomic level, a production function relates the number of employees and machines to the size of the firm's output. For example, the production function for a textile factory would reveal how much more output the factory could produce if it hired (say) 50 additional workers and purchased five more looms. Production functions often are derived from engineering specifications that relate given amounts of physical input to the amount of physical output that can be produced with that input. At the national or economywide level, production functions describe the relationship of the size of a country's total labor force and the value of its capital stock with the level of that country's gross domestic product (GDP; its total output). These economywide relationships are called **aggregate production functions**.

[3]Robert Solow, "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics* 70 (February 1956), 65–94.

[4]This five-equation presentation is based on teaching notes compiled by World Bank economist Shantayanan Devarajan, to whom we are indebted.

Our first equation is an aggregate production function. If Y represents total output (and therefore total income), K is the capital stock, and L is the labor supply, at the most general level, the aggregate production function can be expressed as

$$Y = F(K, L) \qquad [4\text{-}1]$$

This expression indicates that output is a function (denoted by F) of the capital stock and the labor supply. As the capital stock and labor supply grow, output expands. Economic growth occurs by increasing either the capital stock (through new investment in factories, machinery, equipment, roads, and other infrastructure), the size of the labor force, or both. The exact form of the function F (stating precisely *how much* output expands in response to changes in K and L) is what distinguishes many different models of growth, as we will see later in the chapter. The other four equations of the model describe how these increases in K and L come about.

Equations 4–2 through 4–4 are closely linked and together describe how the capital stock (K) changes over time. These three equations first calculate total saving, then relate saving to new investment, and finally describe how new investment changes the size of the capital stock. To calculate saving, we take the most straightforward approach and assume that saving is a fixed share of income:

$$S = sY \qquad [4\text{-}2]$$

In this equation, S represents the total value of saving, and s represents the average saving rate. For example, if the average saving rate is 20 percent and total income is $10 billion, then the value of saving in any year is $2 billion. We assume that the saving rate s is a constant, which for most countries is between 10 and 40 percent (typically averaging between 20 and 25 percent), although for some countries it can be higher or lower. China's savings rate in 2008 (along with those of several large oil exporters) exceeded 50 percent, while several countries (including Mozambique, Guinea, the Seychelles, and Georgia) reported savings rates less than 5 percent of GDP. Actual saving behavior is more complex than this simple model suggests (as we discuss in Chapter 10), but this formulation is sufficient for us to explore the basic relationships between saving, investment, and growth.

The next equation relates total saving (S) to investment (I). In our model, with only one good, there is no international trade (because everyone makes the same product, there is no reason to trade). In a closed economy (one without trade or foreign borrowing), saving must be equal to investment. All output of goods and services produced by the economy must be used for either current consumption or investment, while all income earned by households must be either consumed or saved. Because output is equal to income, it follows that saving must equal investment. This relationship is expressed as

$$S = I \qquad [4\text{-}3]$$

We are now in a position to show how the capital stock changes over time. Two main forces determine changes in the capital stock: new investment (which adds to the capital stock) and depreciation (which slowly erodes the value of the existing capital stock over time). Using the Greek letter delta (Δ) to represent the change in the value of a variable, we express the change in the *capital* stock as ΔK, which is determined as follows:

$$\Delta K = I - (dK) \qquad [4\text{-}4]$$

In this expression d is the rate of depreciation. The first term (I) indicates that the capital stock *increases* each year by the amount of new investment. The second term $-(d \times K)$ shows that the capital stock *decreases* every year because of the depreciation of existing capital. We assume here that the depreciation rate is a constant, usually in the range of 2 to 10 percent.

To see how this works, let us continue our earlier example, in which total income is \$10 billion and saving (and therefore investment) is \$2 billion. Say that the value of the existing capital stock is \$30 billion and the annual rate of depreciation is 3 percent. In this example, the capital stock increases by \$2 billion because of new investment but also decreases by \$0.9 billion (3 percent $\times$ \$30 billion) because of depreciation. Equation 4-4 puts together these two effects, calculating the change in the capital stock as $\Delta K = I - (d \times K) = \$2\text{ billion} - (0.03 \times \$30\text{ billion}) = \$1.1\text{ billion}$. Thus the capital stock increases from \$30 billion to \$31.1 billion. This new value of the capital stock then is inserted into the production function in equation 4-1, allowing for the calculation of a new level of output, Y.

The fifth and final equation of the model focuses on the supply of labor. To keep things simple, we assume that the labor force grows exactly as fast as the total population. Over long periods of time, this assumption is fairly accurate. If n is equal to the growth rate of both the population and the labor force, then the change in the labor force (ΔL) is represented by

$$\Delta L = nL \qquad [4\text{-}5]$$

If the labor force consists of 1 million people and the population (and labor force) is growing by 2 percent, the labor force increases annually by 20,000 (1 million $\times$ 0.02) workers. The labor force now consists of 1.02 million people, a figure that can be inserted into the production function for L to calculate the new level of output. (If we divide both sides of equation 4-5 by L, we can see directly the rate of growth of the labor force, $\Delta L/L = n$.)

These five equations represent the complete model.[5] Collectively, they can be used to examine how changes in population, saving, and investment initially affect

[5]Note that because the model has five equations and five variables (Y, K, L, I, and S) it always can be solved. In addition, there are three fixed parameters (d, s, and n), the values of which are assumed to be fixed exogenously, or outside the system.

the capital stock and labor supply and ultimately determine economic output. New saving generates additional investment, which adds to the capital stock and allows for increased output. New workers add further to the economy's capacity to increase production.

One way these five equations can be simplified slightly is to combine equations 4-2, 4-3, and 4-4. The aggregate level of saving (in equation 4-2) determines the level of investment in equation 4-3, which (together with depreciation) determines changes in the capital stock in equation 4-4. Combining these three equations gives us

$$\Delta K = sY - dK \qquad [4\text{-}6]$$

This equation states that the change in the capital stock (ΔK) is equal to saving (sY) minus depreciation (dK). This expression allows us to calculate the change in the capital stock and enter the new value directly into the aggregate production function in equation 4-1.

THE HARROD-DOMAR GROWTH MODEL

As we have stressed, the aggregate production function (shown earlier as equation 4-1) is at the heart of every model of economic growth. This function can take many different forms, depending on what we believe is the true relationship between the factors of production (K and L) and aggregate output. This relationship depends on (among other things) the mix of economic activities (for example, agriculture, heavy industry, light labor-intensive manufacturing, high-technology processes, services), the level of technology, and other factors. Indeed, much of the theoretical debate in the academic literature on economic growth is about how to best represent the aggregate production process.

THE FIXED-COEFFICIENT PRODUCTION FUNCTION

One special type of a simple production function is shown in Figure 4-1. Output in this figure is represented by **isoquants**, which are combinations of the inputs (labor and capital in this case) that produce equal amounts of output. For example, on the first (innermost) isoquant, it takes capital (plant and equipment) of \$10 million and 100 workers to produce 100,000 keyboards per year (point a). Alternatively, on the second isoquant, \$20 million of capital and 200 workers can produce 200,000 keyboards (point b). Only two isoquants are shown in this diagram, but a nearly infinite number of isoquants are possible, each for a different level of output.

The L-shape of the isoquants is characteristic of a particular type of production function known as **fixed-coefficient production functions**. These production

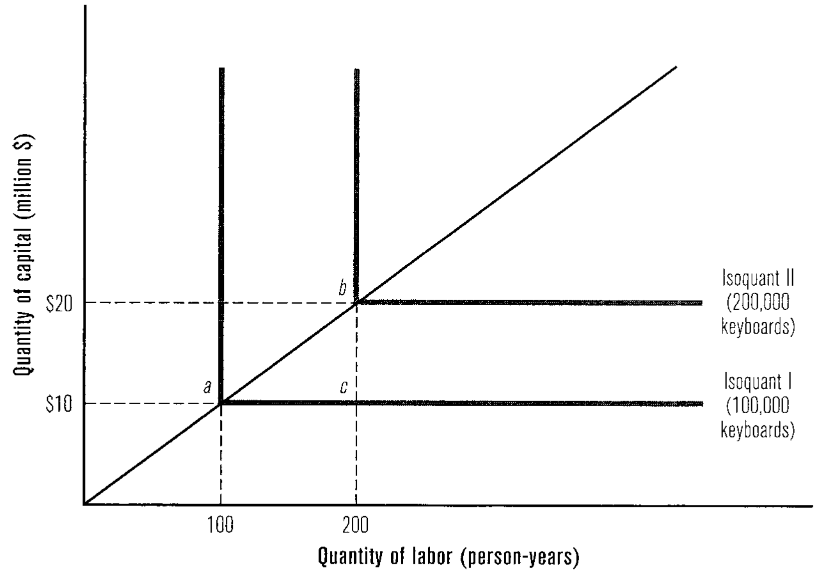

FIGURE 4-1 Isoquants for a Fixed-Coefficient Production Technology
With constant returns to scale, the isoquants will be L-shaped and the production function will be the straight line through their minimum-combination points.

functions are based on the assumption that capital and labor need to be used in a fixed proportion to each other to produce different levels of output. In Figure 4–1, for the first isoquant, the **capital–labor ratio** is 10 million:100, or 100,000:1. In other words, $100,000 in capital must be matched with one worker to produce the given output. For the second isoquant, the ratio is the same: $20 million:200, or 100,000:1. This constant capital–labor ratio is represented in Figure 4–1 by the slope of the ray from the origin through the vertices (*a* and *b*) of the isoquants. These vertices represent the least cost and hence most efficient mix of capital and labor to produce a given quantity of output. In the case of fixed coefficients, this mix is the same for every quantity of output.

With this kind of production function, if more workers are added *without* investing in more capital, output does *not* rise. Look again at the first isoquant, starting at the elbow (with 100 workers and $10 million in capital). If the firm adds more workers (say, increasing to 200 workers) without adding new machines, it moves horizontally to the right along the first isoquant to point *c*. But at this point, or at any other point on this isoquant, the firm still produces just 100,000 keyboards. In this kind of production function, new workers need more machines to increase output. Adding new workers without machines results in idle workers, with no increase in output. Similarly, more machinery without additional workers results in underused machines. On each isoquant, the most efficient production point is at the elbow, where the minimum amounts of capital and labor are used. To use any more of either factor without increasing the other is a waste.

The production technology depicted in Figure 4–1 also is drawn with **constant returns to scale**, so if capital is doubled to $20 million and labor is doubled to 200 workers, output also exactly doubles to 200,000 keyboards per year.[6] With this further assumption, two more ratios remain constant at any level of output: capital to output and labor to output. If keyboards are valued at $50 each, then 100,000 keyboards are worth $5 million. In this case, in the first isoquant, $10 million in capital is needed to produce $5 million worth of keyboards, so the **capital–output ratio** is $10 million: $5 million, or 2:1. In the second isoquant the ratio is the same ($20 million:$10 million, or 2:1). Similarly, for each isoquant the **labor–output ratio** is also a constant, in this case equal to 1:50,000, meaning that each worker produces $50,000 worth of keyboards, or 1,000 keyboards each.

THE CAPITAL–OUTPUT RATIO AND THE HARROD-DOMAR FRAMEWORK

The fixed-coefficient, constant-returns-to-scale production function is the centerpiece of a well-known early model of economic growth that was developed independently during the 1940s by economists Roy Harrod of England and Evsey Domar of MIT, primarily to explain the relationship between growth and unemployment in advanced capitalist societies.[7] It ultimately focuses attention on the role of capital accumulation in the growth process. The **Harrod-Domar model** has been used extensively (perhaps even overused) in developing countries to examine the relationship between growth and capital requirements. The model is based on the real-world observation that some labor is unemployed and proceeds on the basis that capital is the binding constraint on production and growth. In the model, the production function has a very precise form, in which output is assumed to be a *linear* function of capital (and only capital). As usual, the model begins by specifying the level of output, which we later modify to explore changes in output, or economic growth. The production function is specified as follows:

$$Y = 1/v \times K \quad or \quad Y = K/v \qquad [4\text{–}7]$$

where v is a constant. In this equation, the capital stock is multiplied by the fixed number $1/v$ to calculate aggregate production. If $v = 3$ and a firm has $30 million in capital, its annual output would be $10 million. It is difficult to imagine a simpler

[6]In a constant-returns-to-scale production function, if we multiply both capital and labor by any number, w, output multiplies by the same number. In other words, the production function has the following property: $wY = F(wK, wL)$.

[7]Roy F. Harrod, "An Essay in Dynamic Theory," *Economic Journal* (1939), 14–33; Evsey Domar, "Capital Expansion, Rate of Growth, and Employment," *Econometrica* (1946), 137–47; and Domar, "Expansion and Employment," *American Economic Review* 37 (1947), 34–55.

production function. The constant v turns out to be the capital-output ratio because, by rearranging the terms in equation 4-7, we find

$$v = K/Y \qquad [4\text{-}8]$$

The capital-output ratio is a very important parameter in this model, so it is worth dwelling for a moment on its meaning. This ratio essentially is a measure of the productivity of capital or investment. In the earlier example in Figure 4-1, it took $10 million in investment in a new plant and new equipment to produce $5 million worth of keyboards, implying a capital-output ratio of 2:1 (or just 2). A larger v implies that more capital is needed to produce the same amount of output. So, if v were 4 instead, then $20 million in investment would be needed to produce $5 million worth of keyboards.

The capital-output ratio provides an indication of the capital intensity of the production process. In the basic growth model, this ratio varies across countries for two reasons: either the countries use different technologies to produce the same goods or they produce a different mix of goods. Where farmers produce maize using tractors, the capital-output ratio will be much higher than in countries where farmers rely on a large number of workers using hoes and other hand tools. In countries that produce a larger share of **capital-intensive products** (that is, those that require relatively more machinery, such as automobiles, petrochemicals, and steel), v is higher than in countries producing more **labor-intensive products** (such as textiles, basic agriculture, and foot wear). In practice, as economists move from the v of the model to actually measuring it in the real world, the observed capital-output ratio can also vary for a third reason: differences in efficiency. A larger measured v can indicate less-efficient production when capital is not being used as productively as possible. A factory with lots of idle machinery and poorly organized production processes has a higher capital-output ratio than a more-efficiently managed factory.

Economists often calculate the **incremental capital-output ratio (ICOR)** to determine the impact on output of additional (or incremental) capital. The ICOR measures the productivity of additional capital, whereas the (average) capital-output ratio refers to the relationship between a country's total stock of capital and its total national product. In the Harrod-Domar model, because the capital-output ratio is assumed to remain constant, the average capital-output ratio is equal to the incremental capital-output ratio, so the ICOR $= v$.

So far, we have been discussing total output, not growth in output. The production function in equation 4-7 easily can be converted to relate *changes* in output to *changes* in the capital stock:

$$\Delta Y = \Delta K/v \qquad [4\text{-}9]$$

The growth rate of output, g, is simply the increment in output divided by the total amount of output, $\Delta Y/Y$. If we divide both sides of equation 4-9 by Y, then

$$g = \Delta Y/Y = \Delta K/Yv \qquad [4\text{-}10]$$

Finally, from equation 4-6, we know that the change in the capital stock ΔK is equal to saving minus the depreciation of capital ($\Delta K = sY - dK$). Substituting the right-hand side of equation 4-6 into the term for ΔK in equation 4-10 and simplifying[8] leads to the basic Harrod-Domar relationship for an economy:

$$g = (s/v) - d \quad [4\text{-}11]$$

Underlying this equation is the view that capital created by investment is the main determinant of growth in output and that saving makes investment possible.[9] It rivets attention on two keys to the growth process: saving (s) and the productivity of capital (v). The message from this model is clear: Save more and make productive investments, and your economy will grow.

Economic analysts can use this framework either to predict growth or to calculate the amount of saving required to achieve a target growth rate. The first step is to try to estimate the incremental capital-output ratio (v) and depreciation rate (d). With a given saving rate, predicting the growth rate is straightforward. If the saving (or investment) rate is 24 percent, the ICOR is 3, and the depreciation rate is 5 percent, then the economy can be expected to grow by 3 percent (because $0.24/3 - 0.05 = 0.03$).

How does this model work in practice? Consider Indonesia, which from 2002 to 2007 had an investment rate of about 30 percent and recorded a GDP growth just under 5.5 percent per year. Assuming a depreciation rate of 5 percent, the implied ICOR was approximately $v = 2.86$.[10] Would these figures have helped the Indonesian government predict the 2007–08 growth rate? In 2008, the investment rate was 29 percent, so the Harrod-Domar model would have predicted growth of 5.1 percent ($g = 0.29/2.86 - 0.05$). The actual growth rate in 2007–08 was 6.0 percent, within sight of the prediction but not highly accurate.

STRENGTHS AND WEAKNESSES OF THE HARROD-DOMAR FRAMEWORK

The basic strength of the Harrod-Domar model is its simplicity. The data requirements are small, and the equation is easy to use and to estimate. And, as we saw with the example of Indonesia, the model can be somewhat accurate from one year to the next. Generally speaking, in the absence of severe economic shocks (such as a drought, a financial crisis, or large changes in export or import prices), the model can do a reasonable job of estimating expected growth rates in most countries over very short periods of time (a few years). Another strength is its focus on the key role of saving. As discussed

[8]Substituting equation 4-6 into 4-10 leads to $g = (sY - dK)/Y \times 1/v$, which can be simplified to $g = (s - d \times K/Y) \times 1/v$. Since $K/Y = v$, we have $g = (s - dv) \times 1/v$, which leads to $g = s/v - d$.

[9]For an important early contribution to the discussion of the importance of capital accumulation to the growth process, see Joan Robinson, *The Accumulation of Capital* (London: Macmillan, 1956).

[10]Because $g = s/v - d$, then $v = s/(g + d)$. For Indonesia between 2002 and 2007, $v = 0.30/(0.055 + 0.05) = 2.86$.

in Chapter 3, individual decisions about how much income to save and consume are central to the growth process. People prefer to consume sooner rather than later, but the more that is consumed, the less that can be saved to finance investment. The Harrod-Domar model makes it clear that saving is crucial for income to grow over time.

The model, however, has some major weaknesses. One follows directly from the strong focus on saving. Although saving is necessary for growth, the simple form of the model implies that it is also sufficient, which it is not. As pointed out in Chapter 3, the investments financed by saving actually have to pay off with higher income in the future, and not all investments do so. Poor investment decisions, changing government policies, volatile world prices, or simply bad luck can alter the impact of new investment on output and growth. Sustained growth depends both on generating new investment and ensuring that investments are productive over time. In this vein, the allocation of resources across different sectors and firms can be an important determinant of output and growth. Because (for simplicity) the Harrod-Domar assumes only one sector, it leaves out these important allocation issues.

Perhaps the most important limitations in the model stem from the rigid assumptions of fixed capital-to-labor, capital-to-output, and labor-to-output ratios, which imply very little flexibility in the economy over time. To keep these ratios constant, capital, labor, and output must all grow at exactly the same rate, which is highly unlikely to happen in real economies. To see why these growth rates must all be the same, consider the growth rate of capital. If the capital stock grew any faster or slower than output at rate g, the capital-output ratio would change. Thus the capital stock must grow at g to keep the capital-output ratio constant over time. With respect to labor, in our original five-equation model, we stipulated (in equation 4–5) that the labor force would grow at exactly the same pace as the population at rate n. Therefore, the only way that the capital stock and the labor force can grow at the same rate is if n happens to be equal to g. This happens only when $n = g = s/v - d$, and there is no particular reason to believe the population will grow at that rate.

In this model, the economy remains in equilibrium with full employment of the labor force and the capital stock *only* under the very special circumstances that labor, capital, and output all *grow* at the rate g. On the one hand, if n is larger than g, the labor force grows faster than the capital stock. In essence, the saving rate is not high enough to support investment in new machinery sufficient to employ all new workers. A growing number of workers do not have jobs and unemployment rises indefinitely. On the other hand, if g (or $s/v - d$) is larger than n, the capital stock grows faster than the workforce. There are not enough workers for all the available machines, and capital becomes idle. The actual growth rate of the economy no longer is g, as the model stipulates, but slows to n, with output constrained by the number of available workers.

So, unless $s/v - d$ (or g) is exactly equal to n, either labor or capital is not fully employed and the economy is not in a stable equilibrium. This characteristic of the Harrod-Domar model has come to be known as the *knife-edge* problem. As long as

$g = n$, the economy remains in equilibrium, but as soon as either the capital stock or the labor force grows faster than the other, the economy falls off the edge with continuously growing unemployment of either capital or labor.

The rigid assumptions of fixed capital-output, labor-output, and capital-labor ratios may be reasonably accurate for short periods of time or in very special circumstances but almost always are inaccurate over time as an economy evolves and develops. Each of these varies among countries and, for a single country, over time. Consider the incremental capital-output ratio. The productivity of capital can change in response to policy changes, which in turn affects v. Moreover, the capital intensity of the production process can and usually does change over time. A poor country with a low saving rate and surplus labor (unemployed and underemployed workers) can achieve higher growth rates by utilizing as much labor as possible and thus relatively less capital. For example, a country relying heavily on labor-intensive agricultural production will record a low v. As economies grow and per capita income rises, the labor surplus diminishes and economies shift gradually toward more capital-intensive production. As a result, the ICOR shifts upward. Thus a higher v may not necessarily imply inefficiency or slower growth. ICORs can also shift through market mechanisms, as prices of labor and capital change in response to changes in supplies. As growth takes place, saving tends to become relatively more abundant and hence the price of capital falls while employment and wages rise. Therefore, all producers increasingly economize on labor and use more capital and the ICOR tends to rise.

Consider again the example of Indonesia. The ICOR changed from approximately 2.4 during the 1980s, to 4.1 during the 1990s, to 3.6 during 2000–09, reflecting a trend toward more capital-intensive production processes. To continue to use the 1980s ICOR in 2009 would have been very misleading and betrayed a significant misunderstanding of the growth process. The structure of the economy had changed substantially during that time period and, with it, the ICOR. Thailand provides a similar example, as described in Box 4–1.

As a result of these rigidities, the Harrod-Domar framework tends to become increasingly inaccurate over longer periods of time as the actual ICOR changes and, with it, the capital-labor ratio. In a world with fixed-coefficient production functions, little room is left for a factory manager to increase output by hiring one more worker without buying a machine to go with the worker or to purchase more machines for the current workforce to use. The fixed-proportion production function does not allow for any substitution between capital and labor in the production process. In the real world, of course, at least some substitution between labor and capital is possible in most production processes. As we see in the next section, adding this feature to the model allows for a much richer exploration of the growth process.

A final weakness of the Harrod-Domar model is the absence of any role for productivity growth—the ability to produce increasing quantities of output per unit of input. In Figure 4–1, increased factor productivity can be represented by an inward shift of each isoquant toward the origin, implying that less labor and capital would

BOX 4-1 ECONOMIC GROWTH IN THAILAND

In the 1960s, Thailand's agrarian economy depended heavily on rice, maize, rubber, and other agricultural products. About three-quarters of the Thai population derived its income from agricultural activities. Gross domestic product (GDP) per capita in 1960 (measured in 2005 international dollars) was around $1,200—less than one-tenth the average income in the United States. Life expectancy was 53 years and the infant mortality rate was 103 per 1,000 births. Few observers expected Thailand to develop rapidly.

However, since the mid-1960s, the Thai economy has grown rapidly (if not always steadily), benefiting from relatively sound economic management and a favorable external environment. The government regularly achieved surpluses on the current account of its budget and used these funds (plus modest inflows of foreign assistance) to finance investments in rural roads, irrigation, power, telecommunications, and other basic infrastructure. At least until the mid-1990s, the government's fiscal, monetary, and exchange rate policies kept the macroeconomy relatively stable with fairly low inflation, despite the turbulent period of world oil price shocks in the 1970s and 1980s. Beginning in the 1970s, the government began to remove trade restrictions and promote the production of labor-intensive manufactured exports. These products found a ready market in the booming Japanese economy of the 1980s and provided a growing number of jobs for Thai workers.

Thailand's ability to make investments and deepen its capital stock depended on its capacity to save. The country's saving rate averaged about 20 percent in the 1960s, already high for developing countries, and increased steadily over time to an average of 35 percent in the 1990s, falling to about 32 percent during the 2000s. These high saving rates, combined with relatively prudent economic policies, supported very rapid economic growth and development.

Thailand's development experience has been far from completely smooth, however. In mid-1997, a major financial crisis erupted. Huge short-term offshore borrowing combined with a fixed exchange rate and weak financial institutions led to a collapse of a real estate bubble, rapid capital flight, a substantial depreciation of the Thai baht, and a deep recession (see Chapter 13). In some ways, Thailand had become the victim of its own success, with its rapid growth attracting significant numbers of investors looking to gain quick profits, who rapidly fled once the bubble began to collapse. After two years of negative growth (with GDP falling 10 percent in 1998), the economy began to recover and growth rebounded to 3.9 percent between 1999 and 2007.

Over the longer period between 1960 and 2007, per capita growth averaged 4.5 percent, so that the average income in Thailand is now more than eight times higher than it was in 1960. By 2009, life expectancy grew to 69 years, infant mortality fell to 12 per thousand, and adult literacy reached 93 percent. During this period, the structure of the economy changed significantly. By 2009, manufacturing accounted for 34 percent of GDP, up from just 14 percent in 1965, while the share of agricultural production dropped commensurately. The composition of exports shifted away from rice, maize, and other agricultural commodities toward labor-intensive manufactured products, which now account for a large majority of all exports. As the Harrod-Domar model predicts, Thailand's high saving rate and resulting capital accumulation was accompanied by a dramatic increase in output (and income) per capita. Contrary to the Harrod-Domar model, however, the ICOR did not remain constant. As the stock of capital grew and the economy shifted toward more capital-intensive production techniques, the ICOR increased from 2.6 in the 1970s to nearly 5 by the early 2000s. The rising ICOR indicated that, as the Thai economy expanded and the level of capital per worker increased, an ever-larger increment of new capital was required to bring about a given increase in total output.

be needed to produce the same amount of output. The simplest way to capture this in the Harrod-Domar framework is to introduce a smaller ICOR, but of course, this would contradict the idea of a constant ICOR.

Despite these weaknesses, the Harrod-Domar model is still used to a surprisingly wide extent. Economist William Easterly documented how the World Bank and other institutions use the model to calculate "financing gaps" between the amount of available saving and the amount of investment supposedly needed to achieve a target growth rate.[11] He shows how simplistic and sometimes careless use of the model can lead to weak analysis and faulty conclusions. In essence, analysts enamored by the simplicity of the model tend to overlook its shortcomings when applying it to the real world.

The Harrod-Domar model provides some useful insights but does not take us very far. The fixed-coefficient assumption provides the model with very little flexibility and does not capture the ability of real world firms to change the mix of inputs in the production process. The model can be reasonably accurate from one year to the next (in the absence of shocks), and it rightly focuses attention on the importance

[11]See William Easterly, "Aid for Investment," *The Elusive Quest for Growth* (Cambridge: MIT Press, 2001), chap. 2; and Easterly, "The Ghost of the Financing Gap: Testing the Growth Model of the International Financial Institutions," *Journal of Development Economics* 60, no. 2 (December 1999), 423–38.

of saving. But it is quite inaccurate for most countries over longer periods of time and implies that saving is sufficient for growth, although it is not. Indeed, in the late 1950s, Domar expressed strong doubts about his own model, pointing out that it was originally designed to explore employment issues in advanced economies rather than growth per se and was too rigid to be useful for explaining long-term growth.[12] Instead, he endorsed the new growth model of Robert Solow, to which we now turn our attention.

THE SOLOW (NEOCLASSICAL) GROWTH MODEL

THE NEOCLASSICAL PRODUCTION FUNCTION

In 1956, MIT-economist Robert Solow introduced a new model of economic growth that was a big step forward from the Harrod-Domar framework.[13] Solow recognized the problems that arose from the rigid production function in the Harrod-Domar model. Solow's answer was to drop the fixed-coefficients production function and replace it with a **neoclassical production function** that allows for more flexibility and substitution between the factors of production. In the Solow model, the capital–output and capital–labor ratios no longer are fixed but vary, depending on the relative endowments of capital and labor in the economy and the production process. Like the Harrod-Domar model, the Solow model was developed to analyze industrialized economies, but it has been used extensively to explore economic growth in all countries around the world, including developing countries. The Solow model has been enormously influential and remains at the core of most theories of economic growth in developing countries.

The isoquants that underlie the neoclassical production function are shown in Figure 4–2. Note that the isoquants are curved rather than L-shaped as in the fixed-coefficient model. In this figure, at point *a*, $10 million of capital and 100 workers combine to produce 100,000 keyboards, which would be valued at $5 million (because, as stated earlier, keyboards are priced at $50 each). Starting from this point, output could be expanded in any of three ways. If the firm's managers decided to expand at constant factor proportions and move to point *b* on isoquant II to produce 200,000 keyboards, the situation would be identical to the fixed propor-

[12]Evsey Domar, *Essays in the Theory of Economic Growth* (Oxford: Oxford University Press, 1957).

[13]The two classic references of Solow's work are his "A Contribution to the Theory of Economic Growth" and "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics* 39 (August 1957), 312–20. For an excellent and very thorough undergraduate exposition of the Solow and other models of economic growth, see Charles I. Jones, *Introduction to Economic Growth* (New York: W. W. Norton and Company, 2001). In 1987, Solow was awarded the Nobel Prize in economics, primarily for his work on growth theory.

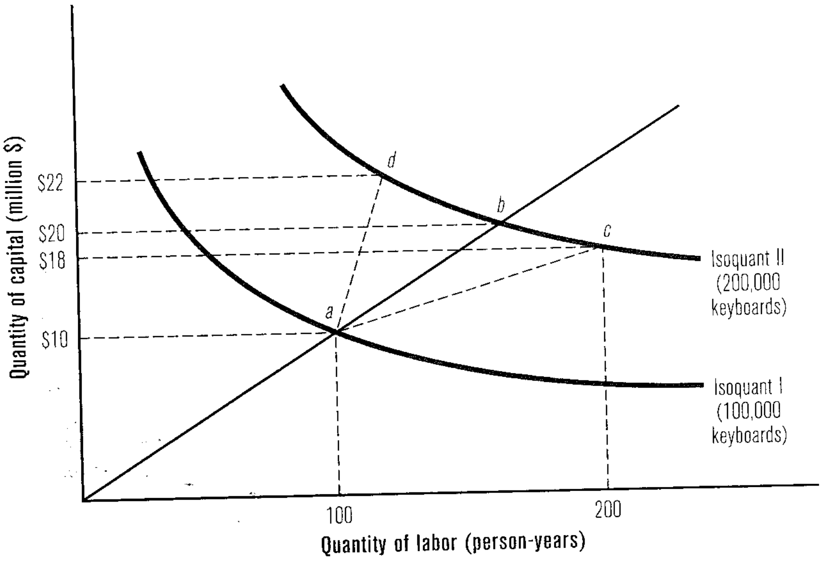

FIGURE 4-2 Isoquants for a Neoclassical Production Technology
Instead of requiring fixed factor proportions, as in Figure 4-1, output can be achieved with varying combinations of labor and capital. This is called a *neoclassical* production function. Note that the isoquants are curved rather than L-shaped.

tions case of Figure 4-1. The capital-output ratio at both points *a* and *b* would be 2:1, as it was before ($10 million of capital produces $5 million of keyboards at point *a*, and $20 million of capital produces $10 million of keyboards at point *b*). Note that the Solow model retains from the Harrod-Domar model the assumption of constant returns to scale, so that a doubling of labor and capital leads to a doubling of output. But by dropping the fixed-coefficients assumption, production of 200,000 keyboards could be achieved by using different combinations of capital and labor. For example, the firm could use more labor and less capital (a more labor-intensive method), such as at point *c* on isoquant II. In that case, the capital-output ratio falls to 1.8:1 ($18 million in capital to produce $10 million in keyboards).

Alternatively, the firm could choose a more capital-intensive method, such as at point *d* on isoquant II, where the capital-output ratio would rise to 2.2:1 ($22 million in capital to produce $10 million in keyboards). The kinds of tools that policy makers use to try to decrease or increase the capital-output ratio are discussed in depth in several chapters later in this text.

THE BASIC EQUATIONS OF THE SOLOW MODEL

The Solow model is understood most easily by expressing all the key variables in per-worker terms (for example, output per worker and capital per worker). To do so, we divide both sides of the production function in equation 4-1 by L, so that it takes the form

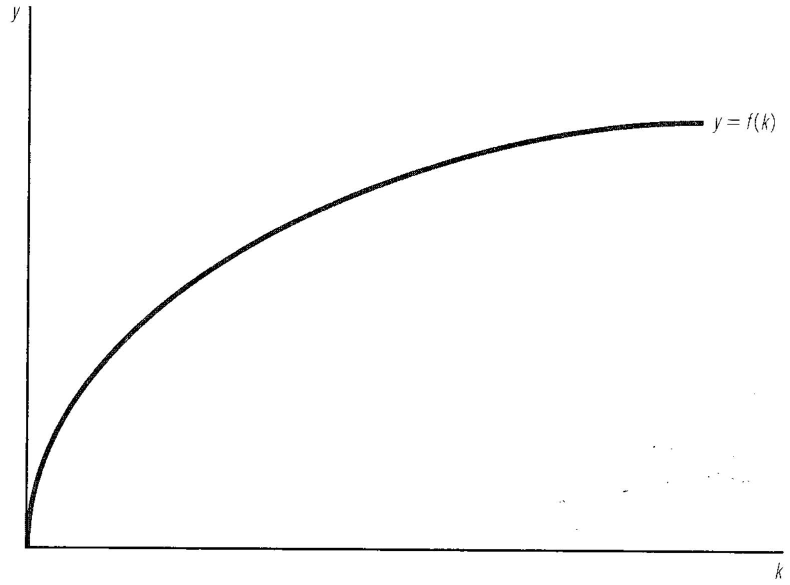

FIGURE 4-3 The Production Function in the Solow Growth Model
The neoclassical production function in the Solow model displays diminishing returns to capital so that each additional increment in capital per worker (k) is associated with smaller increases in output per worker (y).

$$Y/L = F(K/L, 1) \qquad [4\text{-}12]$$

The equation shows that output per worker is a function of capital per worker.[14] If we use lower-case letters to represent quantities in per-worker terms, then y is output per worker (that is, $y = Y/L$) and k is capital per worker ($k = K/L$). This gives us the first equation of the Solow model, in which the production function can be written simply as

$$y = f(k) \qquad [4\text{-}13]$$

Solow's model assumes a production function with the familiar property of **diminishing returns to capital**. With a fixed labor supply, giving workers an initial amount of machinery to work with results in large gains in output. But as these workers are given more and more machinery, the addition to output from each new machine gets smaller and smaller. An aggregate production function with this property is shown in Figure 4-3. The horizontal axis represents capital per worker (k), and the vertical axis shows output per worker (y). The slope of the curve declines as the capital stock increases, reflecting the assumption of the diminishing marginal product of capital.

[14] We can divide both sides by L because the Solow model (like the Harrod-Domar model) assumes the production function exhibits constant returns to scale and has the property that $wY = F(wK, wL)$. To express the Solow model in per-worker terms, we let $w = 1/L$.

Each movement to the right on the horizontal axis yields a smaller and smaller increase in output per worker. (By comparison, the fixed-coefficient production function assumed in the Harrod-Domar model would be a straight line drawn through the origin.)

The first equation of the Solow model tells us that capital per worker is fundamental to the growth process. In turn, the second equation focuses on the determinants of changes in capital per worker. This second equation can be derived from equation 4–6[15] and shows that capital accumulation depends on saving, the growth rate of the labor force, and depreciation:

$$\Delta k = sy - (n + d)k \qquad [4\text{–}14]$$

This is a very important equation, so we should understand exactly what it means. It states that the change in capital per worker (Δk) is determined by three things:

- *The Δk is positively related to saving per worker.* Because s is the saving rate and y is income (or output) per worker, the term sy is equal to saving per worker. As saving per worker increases, so does investment per worker, and the capital stock per worker (k) grows.
- *The Δk is negatively related to population growth.* This is shown by the term $-nk$. Each year, because of growth in the population and labor force, there are nL new workers. If there were no new investment, the increase in the labor force would mean that capital *per worker* (k) falls. Equation 4–14 states that capital per worker falls by exactly nk.
- *Depreciation erodes the capital stock.* Each year, the amount of capital per worker falls by the amount $-dk$ simply because of depreciation.

Therefore, saving (and investment) adds to capital per worker, whereas labor force growth and depreciation reduce capital per worker. When saving per capita, sy,

[15]To derive equation 4–14, we begin by dividing both sides of equation 4–6 by K so that

$$\Delta K/K = sY/K - d$$

We then focus on the capital per worker ratio, $k = K/L$. The growth rate of k is equal to the growth rate of K minus the growth rate of L:

$$\Delta k/k = \Delta K/K - \Delta L/L$$

With a little rearranging of terms, this equation can be written as $\Delta K/K = \Delta k/k + \Delta L/L$. We earlier assumed that both the population and the labor force were growing at rate n, so $\Delta L/L = n$. By substitution we obtain

$$\Delta K/K = \Delta k/k + n$$

Note that in both the first equation of this footnote and the equation just given, the left-hand side is equal to $\Delta K/K$. This implies that the right-hand sides of these two equations are equal to each other, as follows:

$$\Delta k/k + n = sY/K - d$$

By subtracting n from both sides and multiplying through by k, we find that

$$\Delta k = sy - nk - dk \quad \text{or} \quad \Delta k = sy - (n + d)k$$

is larger than the amount of new capital needed to compensate for labor force growth and depreciation, $(n + d)k$, then Δk is a positive number. This implies that capital per worker k increases.

The process through which the economy increases the amount of capital per worker, k, is called **capital deepening**. Economies in which workers have access to more machines, computers, trucks, and other equipment have a deeper capital base than economies with less machinery, and these economies are able to produce more output per worker.

In some economies, however, the amount of saving is just enough to provide the same amount of capital to new workers and compensate for depreciation. An increase in the capital stock that just keeps pace with the expanding labor force and depreciation is called **capital widening** (referring to a widening of both the total amount of capital and the size of the workforce). Capital widening occurs when sy is exactly equal to $(n + d)k$, implying no change in k. Using this terminology, equation 4-14 can be restated as saying that *capital deepening* (Δk) *is equal to saving per worker* (sy) *minus the amount needed for capital widening* $[(n + d)k]$.

A country with a high saving rate can easily deepen its capital base and rapidly expand the amount of capital per worker, thus providing the basis for growth in output. In Singapore, for example, where the saving rate has averaged more than 40 percent since the early 1980s, it is not difficult to provide capital to the growing labor force and make up for depreciation and still have plenty left over to supply existing workers with additional capital. By contrast, Kenya, with a saving rate of about 15 percent, has much less saving to spare for capital deepening after providing machines to new workers and making up for depreciation. As a result, capital per worker does not grow as quickly, and neither does output (or income) per worker. Partly because of this large difference in saving rates, output per person in Singapore grew by an average of 4.9 percent per year between 1960 and 2009, while Kenya's growth averaged about 0.34 percent.

We can summarize the two basic equations of the Solow model as follows. The first $[y = f(k)]$ simply states that output per worker (or income per capita) depends on the amount of capital per worker. The second equation, $\Delta k = sy - (n + d)k$, says that change in capital per worker depends on saving, the population growth rate, and depreciation. Thus, as in the Harrod-Domar model, saving plays a central role in the Solow model. However, the relationship between saving and growth is not linear because of diminishing returns to capital in the production function. In addition, the Solow model introduces a role for the population growth rate and allows for substitution between capital and labor in the growth process.

Now that we are equipped with the basic model, we can proceed to analyze the effects of changes in the saving rate, population growth, and depreciation on economic output and economic growth. This is accomplished most easily by examining the model in graphical form.

THE SOLOW DIAGRAM

The diagram of the Solow model consists of three curves, shown in Figure 4-4. The first is the production function $y = f(k)$, given by equation 4-13. The second is a saving function, which is derived directly from the production function. The new curve shows saving per capita, *sy*, calculated by multiplying both sides of equation 4-13 by the saving rate, so that $sy = s \times f(k)$. Because saving is assumed to be a fixed fraction of income (with *s* between 0 and 1), the saving function has the same shape as the production function but is shifted downward by the factor *s*. The third curve is the line $(n + d)k$, which is a straight line through the origin with the slope $(n + d)$. This line represents the amount of new capital needed as a result of growth in the labor force and depreciation just to keep capital per worker (k) constant. Note that the second and third curves are representations of the two right-hand terms of equation 4-14.

The second and third curves intersect at point *A*, where $k = k_0$. (Note that, on the production function above the *sy* curve, $k = k_0$ corresponds to a point directly above *A* where $y = y_0$ on the vertical axis.) At point *A*, *sy* is exactly equal to $(n + d)k$, so capital per worker does not change and k remains constant. At other points along the horizontal axis, the *vertical difference* between the *sy* curve and the $(n + d)k$ line determines the *change* in capital per worker. To the left of point *A* (say, where $k = k_1$

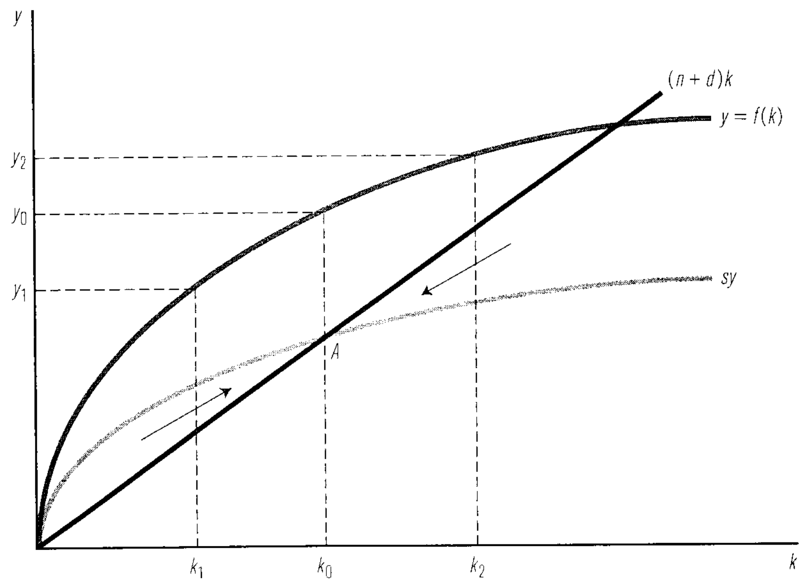

FIGURE 4-4 The Basic Solow Growth Model Diagram

In the basic Solow diagram, point *A* is the only place where the amount of new saving, *sy*, is exactly equal to the amount of new capital needed to compensate for growth in the workforce and depreciation $(n + d)k$. Point *A* is the steady state level of capital per worker and output per worker.

and on the production function $y = y_1$), the amount of saving in the economy per person (sy) is larger than the amount of saving needed to compensate for new workers and depreciation [$(n + d)k$]. As a result, the amount of capital per person (k) grows (capital deepening) and the economy shifts to the right along the horizontal axis. The economy continues to shift to the right as long as the sy curve is *above* the $(n + d)k$ curve, until eventually the economy reaches an equilibrium at point A. In terms of the production function, the shift to the right implies an increase in output per worker, y (or income per capita), from y_1 to y_0. To the right of point A (say, where $k = k_2$ and $y = y_2$), saving per capita is smaller than the amount needed for new workers and depreciation, so capital per worker falls and the economy shifts to the left along the horizontal axis. Once again, this shift continues until the economy reaches point A. The shift to the left corresponds to a decline in output per worker from y_2 to y_0.

Point A is the only place where the amount of new saving, sy, is exactly equal to the amount of new capital needed for growth in the workforce and depreciation. Therefore, at this point, the amount of capital per worker, k, remains constant. Saving per worker (on the vertical axis of the saving function) also remains constant, as does output per worker (or income per capita) on the production function, with $y = y_0$. As a result, point A is called the **steady state** of the Solow model. Output per capita at the steady state (y_0) is alternatively referred to as the **steady state, long run**, or **potential level of output per worker**.

It is very important to note, however, that all the values that remain constant are expressed as *per worker*. Although output per worker is constant, *total* output continues to grow at rate n, the same rate the population and workforce grow. In other words, *at the steady state GDP (Y) grows at the rate n, but GDP per capita (y) is constant (average income remains unchanged)*. Similarly, although capital per worker and saving per worker are constant at point A, total capital and total saving grow.

CHANGES IN THE SAVING RATE AND POPULATION GROWTH RATE IN THE SOLOW MODEL

Both the Solow and Harrod-Domar models put saving (and investment) at the core of the growth process. In the Harrod-Domar model, an increase in the saving rate translates directly (and linearly) into an increase in aggregate output. What is the impact of a higher saving rate in the Solow model?

As shown in Figure 4-5, increasing the saving rate from s to s' shifts the saving function sy up to $s'y$, without shifting either the production function or the capital widening line $(n + d)k$. The increase in the saving rate means that saving per worker (and investment per worker) now is greater than $(n + d)k$, so k gradually increases. The economy shifts to a new long-run equilibrium at point B. In the process, capital per worker increases from k_0 to k_3 and output per worker increases from y_0 to y_3. The aggregate economy initially grows at a rate faster than its steady-state growth

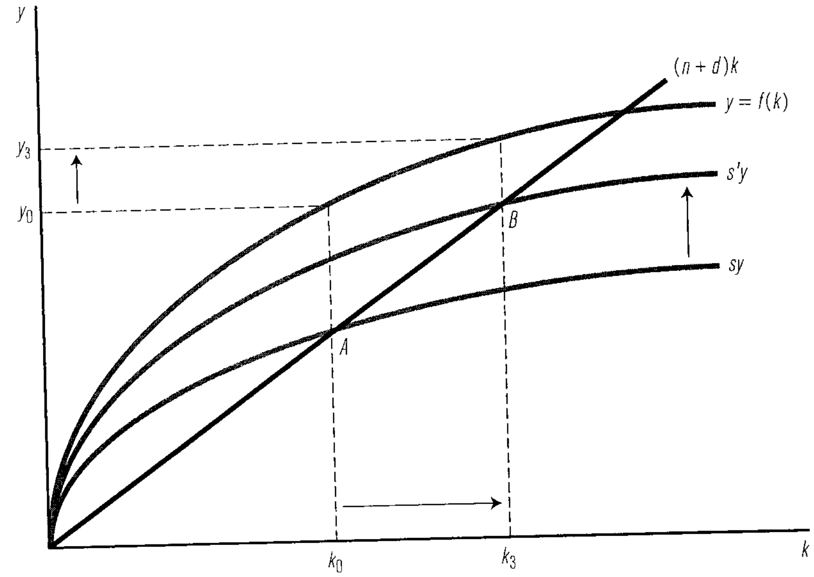

FIGURE 4-5 An Increase in the Saving Rate in the Solow Model
An increase in the saving rate from s to s' results in an upward shift in the capital deepening curve, so that capital per worker increases from $k_0$to k_3.

rate of n until it reaches point B, where the long-run growth rate reverts to n. Thus, the higher saving rate leads to more investment, a permanently higher stock of capital per worker, and a permanently higher level of income (or output) per worker. In other words, the Solow model predicts that economies that save more have higher standards of living than those that save less. (The increase in per capita income, however, is smaller than for a similar increase in s in the Harrod-Domar model because the Solow model has diminishing returns in production.) Higher saving also leads to a *temporary* increase in the economic growth rate as the steady state shifts from A to B. However, the increase in the saving rate does *not* result in a permanent increase in the long-run rate of output growth, which remains at n.

The Solow diagram also can be used to evaluate the impact of a change in the population (or labor force) growth rate. An increase in the population growth rate from n to n' rotates the capital widening line to the left from $(n + d)k$ to $(n' + d)k$, as shown in Figure 4-6. The production and saving functions do not change. Because there are more workers, savings per worker (sy) becomes smaller and no longer is large enough to keep capital per worker constant. Therefore, k begins to decline, and the economy moves to a new steady state, C. More workers also means that capital per worker declines from $k_0$to $k_4$and saving per worker falls from sy_0 to sy_4. Output per worker (or income per capita) also declines, from y_0 to y_4. Thus, an increase in the population growth rate leads to lower average income in the Solow model. Note, however, that the new steady-state growth rate of the entire economy has increased

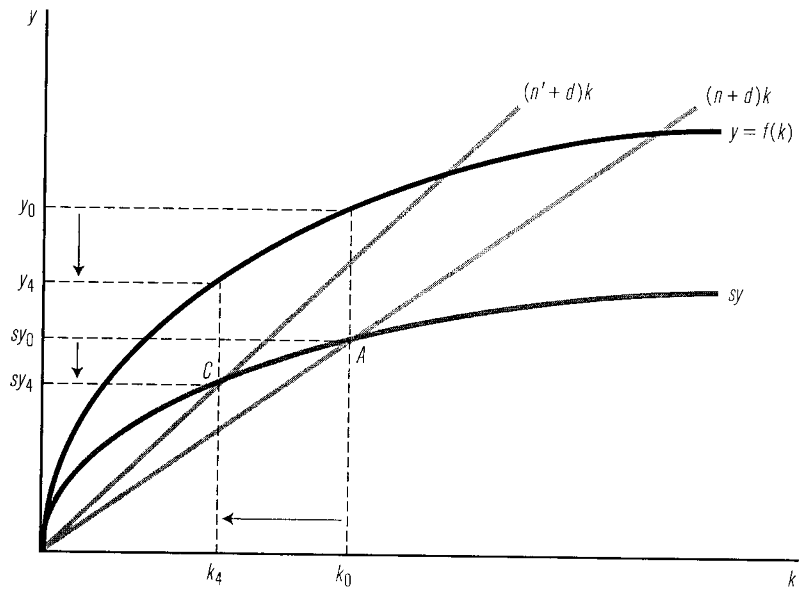

FIGURE 4-6 Changes in the Population Growth Rate in the Solow Model
An increase in the rate of population growth from n to n' causes the capital widening curve to rotate to the left. Equilibrium capital per worker drops from k_0 to k_4.

from n to n' at point *C*. In other words, with a higher population growth rate, Y needs to grow faster to keep y constant.[16] By contrast, a *reduction* in the population growth rate rotates the $(n + d)k$ line to the right and leads to a process of capital deepening, with an increase in both k and in the steady-state level of income per worker, y. However, the relationship between population growth and economic growth is not quite so simple, as described in Box 4-2.

The Solow growth model (as described to this point) suggests that growth rates differ across countries for two main reasons:

- Two countries with the same current level of income may experience different growth rates *if one has a higher steady-state level of income than the other.* To the extent that two countries with the same current level of income have different aggregate production functions, saving rates, population growth rates, or rates of change in productivity (described later), their steady-state income levels will differ and so will their growth rates during the transition to their respective steady states.
- Two countries with the same long-run steady-state level of income may have different growth rates *if they are in different points in the transition to*

[16] A similar exercise can be used to determine the impact of an increase in the depreciation rate, d. Such an increase results in a reduction in k and y to a lower steady-state income per capita. The subtle difference between an increase in n and an increase in d is that the latter case does not lead to a change in the long-run growth rate of Y, which remains equal to n.

BOX 4-2 POPULATION GROWTH AND ECONOMIC GROWTH

The inverse correlation between population growth and economic growth suggested by the Solow model has a lot of intuitive appeal. Countries such as Belize, Burkina Faso, Liberia, Niger, and Yemen have some of the world's fastest rates of population growth. They are also among some of the poorest nations in the world. But closer inspection of the Solow model reveals other predictions about how population growth may affect economic growth, and a further examination of empirical trends suggests a much more complex relationship.

Figure 4–6 illustrates that, in the Solow model, an increase in the rate of population growth lowers the steady-state level of income, y. However, y refers to output per *worker*, whereas the more common measure of aggregate economic welfare is output per person, y^*. These two measures of output, of course, are related to one another, as follows:

$$y^* = y \times (N/Pop)$$

where N equals the number of workers, and Pop is the total population. Differences in the level of output per capita, therefore, depend on both the amount of output per worker and the ratio of workers to total population. Growth in per capita output, similarly, depends on growth in output per worker and the growth in the worker-to-population ratio.

The Solow model suggests that more rapid population growth reduces capital deepening and hence reduces growth in output per worker. But the effect of population growth on the ratio of workers to total population is more complex. It depends on the age structure of the population. Because of rapid population growth, most developing countries have a young age structure, with a larger share of younger people than is the case in developed nations that have growing populations of more elderly people. As a result of previously higher population growth rates, many developing nations today are experiencing an increase in their ratio of workers to total population. This positive effect of a changing age structure on per capita incomes, sometimes referred to as a *demographic gift*, can play a positive and large role in determining economic growth rates.

The impact of population growth on economic growth goes beyond its effects on capital widening and a nation's age structure. In the Solow model, saving and technological change are considered exogenous. But population growth can affect these parameters as well. The net effect of population growth on economic growth is therefore an empirical matter. Econometric investigations of the impact of population growth on economic growth generally show no systematic relationship (see Chapter 7). Population growth influences many aspects of economic growth and development, not only those described by the Solow model.

the steady state. For example, consider two countries that are identical in every way except that one has a higher saving rate than the other and, so, initially has a higher steady-state level of income. At the steady states, the country with the higher saving rate has a higher level of output per worker, but both are growing at the rate n. If the country with the lower saving rate suddenly increases its saving to match the other country, its growth rate will be higher than the other country until it catches up at the new steady state. Thus, even though everything is identical in the two countries, their growth rates may differ during the transition to the steady state, which may take many years.

TECHNOLOGICAL CHANGE IN THE SOLOW MODEL

The Solow model, as described to this point, is a powerful tool for analyzing the interrelationships between saving, investment, population growth, output, and economic growth. However, the unsettling conclusion of the basic model is that, once the economy reaches its long-run potential level of income, economic growth simply matches population growth, with no chance for sustained increases in average income. How can the model explain the historical fact reported in Chapter 2 that many of the world's countries have seen steady growth in average incomes since 1820? Solow's answer was technological change.[17] According to this idea, a key reason why France, Germany, the United Kingdom, the United States, and other high-income countries have been able to sustain growth in per capita income over very long periods of time is that technological progress has allowed output per worker to continue to grow.

Technological progress is a key driver, but not the only driver, of productivity growth. Historically, most technological innovation has originated in today's developed countries, where technological progress has played a central role in explaining productivity growth. Technological progress is also important for today's developing countries; yet, most such progress in developing countries is adopted and adapted from developed countries. In developing countries, the productivity-enhancing effects of technological innovation have played a smaller role in driving productivity growth than in developed countries. In the former, productivity growth has also resulted from improvements in physical infrastructure, increased education of the labor force, and improvements in regulatory environments and incentives. Absent these (and related) factors, technology adoption in developing countries tends to be more limited and less effective in enhancing factor productivity. With these important caveats, we proceed to introduce productivity growth into the Solow model using technological progress as a shorthand.

[17]See Solow, "Technical Change and the Aggregate Production Function." For an early discussion about the relationship between capital accumulation and technological progress, see Joan Robinson, *Essays in the Theory of Economic Growth* (London: Macmillan, 1962).

To incorporate an economy's ability to produce more output with the same amount of capital and labor, we slightly modify the original production function and introduce a variable, T, to represent technological progress, as follows:

$$Y = F(K, T \times L) \qquad [4\text{–}15]$$

In this specification, technology is introduced in such a way that it directly enhances the input of labor, as shown by the specification in which L is multiplied by T. This type of technological change is referred to as *labor augmenting.*[18] As technology improves (T rises), the efficiency and productivity of labor increases because the same amount of labor can now produce more output. Increases in T can result from improvements in technology in the scientific sense (new inventions and processes) or in terms of **human capital**, such as improvements in the health, education, or skills of the workforce.[19]

The combined term $T \times L$ is sometimes referred to as the amount of **effective units of labor**. The expression $T \times L$ measures both the amount of labor and its efficiency in the production process. An increase in either T or L increases the amount of effective labor and therefore increases aggregate production. For example, an insurance sales office can increase its effective workforce by either adding new workers or giving each worker a faster computer or better cell phone. An increase in T differs from an increase in L, however, because the rise in aggregate income from new technology does not need to be shared with additional workers. Therefore, *technological change and productivity growth more broadly allow output (and income) per worker to increase.*

The usual assumption is that technology improves at a constant rate, which we denote by the Greek letter theta (θ), so that $\Delta T/T = \theta$. If technology grows at 1 percent per year, then each worker becomes 1 percent more productive each year. With the workforce growing at n, growth in the effective supply of labor is equal to $n + \theta$. If the workforce (and population) grows by 2 percent per year and technology grows by 1 percent per year, the effective supply of labor increases by 3 percent per year.

To show technological change in the Solow diagram, we need to modify our notation. Whereas earlier we expressed y and k in terms of output and capital *per worker*, we now need to express these variables in terms of output and capital *per effective worker*. The change is straightforward. Instead of dividing Y and K by L as

[18]Two other possibilities are *capital-augmenting technological change* [$Y = F(T \times K, L)$], which enhances capital inputs, and *Hicks-neutral technological change* [$Y = F(T \times K, T \times L)$], which enhances both capital and labor input. For our purposes, the specific way in which technology is introduced does not affect the basic conclusions of the model.

[19]Keep in mind, however, that, while these two broad categories of improvements in technology have similar general effects in this aggregate model, their true effects are somewhat different in the real world. Technological change in the mechanical sense or from the spread of a new idea can be shared widely across the workforce and considered a public good. Improvements in human capital, by contrast, are specific to individual workers and are not necessarily widely shared. However, both have the effect of augmenting the supply of labor and increasing total output.

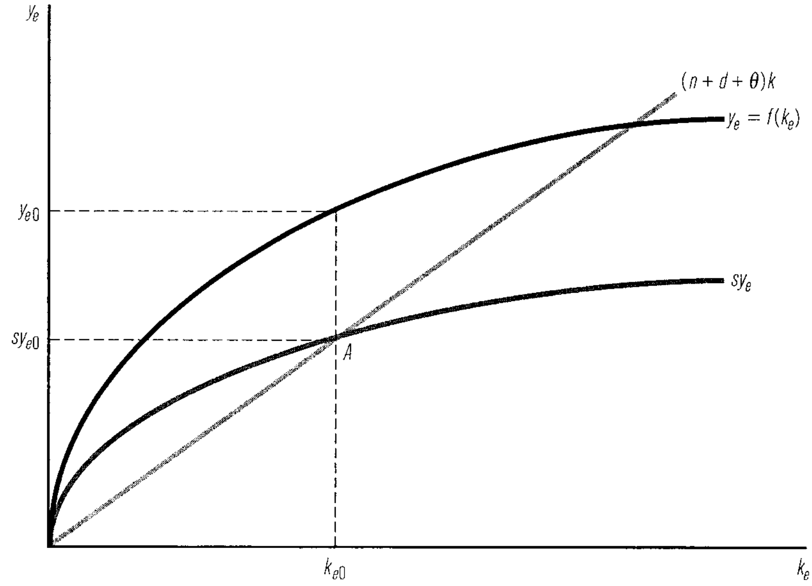

FIGURE 4-7 The Solow Model with Technical Change
In the Solow model with technical change, the equilibrium level of effective capital per worker (K_{e0}) is determined by point A, the intersection of the effective capital widening curve $(n + d + \theta)$ and the effective saving curve (sy_e).

previously (to obtain y and k), we now divide each by $(T \times L)$. Thus **output per effective worker** (y_e) is defined as $y_e = Y/(T \times L)$. Similarly, **capital per effective worker** (k_e) is defined as $k_e = K/(T \times L)$.[20]

With these changes, the production function can be written as $y_e = f(k_e)$ and saving per effective worker expressed as sy_e. With effective labor now growing at the rate $n + \theta$, the capital accumulation equation (4-14) changes to

$$\Delta k_e = sy_e - (n + d + \theta)k_e \qquad [4\text{-}16]$$

The new term $(n + d + \theta)k_e$ is larger than the original $(n + d)k$, indicating that more capital is needed to keep capital *per effective worker* constant.

These changes are shown in Figure 4-7, which looks very similar to the basic Solow diagram, with only a slight change in notation. There still is one steady-state point, at which saving per effective worker is just equal to the amount of new capital needed to compensate for changes in the size of the workforce, depreciation, and technological change in order to keep capital per effective worker constant.

One change, however, is very important. At the steady state, output per *effective* worker is constant, rather than output per worker. *Total output now grows at the rate $n + \theta$, so that output per actual worker (or income per person) increases at rate θ.*

[20]Note that this is consistent with the earlier notation. If there is no technological change (our earlier assumption), so that $T = 1$ (and remains unchanged), then $y_e = y$ and $k_e = k$.

With the introduction of technology, the model now incorporates the possibility of an economy experiencing sustained growth in per capita income at rate θ. This mechanism provides a plausible explanation for why the industrialized countries never seem to reach a steady state with constant output per worker but instead historically have recorded growth in output per worker of between 1 and 2 percent per year.

STRENGTHS AND WEAKNESSES OF THE SOLOW FRAMEWORK

Although the Solow model is more complex than the Harrod-Domar framework, it is a more powerful tool for understanding the growth process. By replacing the fixed coefficients production function with a neoclassical one, the model provides more reasonable flexibility of factor proportions in the production process. Like the Harrod-Domar framework, it emphasizes the important role of factor accumulation and saving, but its assumption of diminishing marginal product of capital provides more realism and accuracy over time. It departs significantly from the Harrod-Domar framework in distinguishing the current level of income per worker from the long-run steady-state level and focuses attention on the transition path to that steady state. The model provides powerful insights into the relationship between saving, investment, population growth, and technological change on the steady-state level of output per worker. The Solow model does a much better job, albeit far from perfect, of describing real world outcomes than the Harrod-Domar model.

A particularly important contribution of the model is the simple yet powerful insights it provides into the role of technological change and productivity growth in the growth process. For policy makers, key questions then become how to best acquire new technologies and how to magnify their potential contributions to productivity growth through complementary investments and policies. For most low-income countries, while some domestic innovation is possible, for many industries it is probably most cost-effective for entrepreneurs to acquire the bulk of their new technologies from other countries (one of the benefits of "globalization") and adapt them to local circumstances. The willingness of entrepreneurs to make such investments, however, may depend on a number of factors that government can influence. Investments in education (for example, improvements in the *quality* of the labor force) may enable firms in developing countries to make better use of imported technologies. Investments in physical infrastructure might motivate technology adoption by better connecting firms to markets. Improved public institutions to protect property rights and extend the rule of law may also provide powerful incentives for firms to invest in improving their own productivity. However, the model's focus on the role of factor accumulation and productivity (including technology) as the proximate determinants of the steady state raises a new set of questions that the model does not answer.

The model's most troubling limitation is that Solow specified productivity growth as exogenous (that is, determined independently of all the variables and parameters specified in the model). He did not spell out exactly how it takes place or how the growth process itself might affect it. In this sense, productivity growth has been called "manna from heaven" in the Solow model. This is an appropriate abstraction for the purpose of explaining the theoretical role of technological change. In practice, policy makers in developing countries need to know the sources of this manna.

What are the more fundamental determinants of factor accumulation and productivity that affect the steady state and the rate of economic growth? The empirical evidence in Chapter 3 suggests that the most rapidly growing developing countries share certain common characteristics: greater economic and political stability, relatively better health and education, stronger governance and institutions, more export oriented trade policies, and more favorable geography. Box 4-3 provides an estimate of the quantitative importance of these factors in East Asia's rapid growth relative to other countries. In the language of the Solow model, these characteristics operate through factor accumulation and productivity to help determine the precise shape of the production function and the steady-state level of output per worker. Changing any of these factors—say, encouraging more open trade—changes the steady-state level of output per worker and therefore the current rate of economic growth as the economy adjusts to the new steady state. Thus, the model helps us focus attention on these more fundamental influences on the steady state and the growth rate, but it does not provide a full understanding of the precise pathways through which these factors influence output and growth.

Certain characteristics of the neoclassical production function embedded in the Solow model also lead to one of the model's broadest and most problematic empirical predictions—that initially poorer countries will grow more rapidly than initially wealthier countries and eventually catch up. To understand this theoretical prediction, we begin by reviewing the neoclassical production function introduced in Chapter 3.

DIMINISHING RETURNS AND THE PRODUCTION FUNCTION

For output and income to continue to grow over time, a country must continue to attract investment and achieve productivity gains. But as the capital stock grows, the magnitude of the impact of new investment on growth may change. Most growth models are based on the assumption that the return on investment declines as the capital stock grows. We illustrate this aspect of the aggregate neoclassical production in Figure 4-8. The shape of this production function reflects the important but common assumption of diminishing returns to capital or, more precisely, a **diminishing marginal product of capital (MPK)**. This property is indicated by the gradual

BOX 4-3 EXPLAINING DIFFERENCES IN GROWTH RATES

Many recent studies have shown that the initial levels of income, openness to trade, healthy populations, effective governance, favorable geography, and high saving rates all contribute to rapid economic growth. But which are most important? One study sought to explain differences in growth during the period 1965–90 among three groups of countries: 10 East and Southeast Asian countries (in which per capita growth averaged 4.6 percent), 17 sub-Saharan African countries (in which growth averaged 0.6 percent), and 21 Latin American countries (in which growth averaged 0.7 percent).[a]

Policy variables explained much of the differences in growth rates. The East and Southeast Asian countries recorded higher government saving rates, were more open to trade, and had higher-quality government institutions. Together, the differences in these policies accounted for 1.7 percentage points of the 4.0 percentage point difference between the East and Southeast Asian and sub-Saharan African growth rates, and 1.8 percentage points of the difference between East and Southeast Asia and Latin America. Openness to trade stood out as the single most important policy choice affecting these growth rates.

Initial levels of income also were important, as the Solow model predicts. Because the Latin American countries had higher average income (and therefore greater output per worker) than the East Asian countries in 1965, the Solow model would predict somewhat slower growth in Latin America. Sure enough, this study estimates that Latin America's higher initial income slowed its growth rate by 1.2 percentage points relative to East and Southeast Asia, after controlling for other factors. By contrast, the sub-Saharan African countries had lower average initial income, indicating that (all else being equal) these countries could have grown 1.0 percentage point faster than the East and Southeast Asian countries, rather than the actual outcome of 4.0 percentage points slower. This suggests that the other factors had to account for a full 5.0 percentage point difference in growth rates between East and Southeast Asia and sub-Saharan Africa.

Initial levels of health, as indicated by life expectancy at birth, were a major factor contributing to sub-Saharan Africa's slow growth. Life expectancy at birth averaged 41 years in sub-Saharan Africa in 1965, compared to 55 years in East

[a]Steven Radelet, Jeffrey Sachs, and Jong-Wha Lee, "The Determinants and Prospects for Economic Growth in Asia," *International Economic Journal* 15, no. 3 (Fall 2001), 1–30. These results are summarized in the Asian Development Bank's study *Emerging Asia: Changes and Challenges* (Manila: Asian Development Bank, 1997), 79–82.

and Southeast Asia. The study estimates that this reduced sub-Saharan Africa's growth rate by 1.3 percentage points relative to East and Southeast Asia. By contrast, because average life expectancy in Latin America in 1965 was almost the same as in East and Southeast Asia, health explains little of the difference in growth between these regions.

Favorable geography helped East and Southeast Asia grow faster. The combination of fewer landlocked countries, longer average coastline, fewer countries located in the deep tropics, and less dependence on natural resource exports all favored Asia. Taken together, these factors accounted for 1.0 percentage point of East and Southeast Asia's rapid growth compared to sub-Saharan Africa, and 0.6 percentage points relative to Latin America. Differences in initial levels of education and the changing demographic structure of the population accounted for the remaining differences in growth rates across these regions.

Of course, this simple accounting framework does not fully explain the complex relationships that underlie economic growth. Each of the variables in the study captures a range of other factors that affect growth rates. For example, differences in government saving rates probably reflect differences in fiscal policy, inflation rates, political stability, and many other factors. Because of lack of sufficient data, the analysis omits several factors (such as environmental degradation) that may be important. And it certainly does not begin to explain why different policy choices were made in different countries. As a result, studies like these should be seen as a first step to understanding growth, rather than as a precise explanation for the many complex differences across countries.

flattening (or declining slope) of the curve as capital per worker grows. It is important, however, to bear in mind the distinction between diminishing returns to individual factors of production holding constant the use of all other factors and the **returns to scale** of the entire production function. The production functions illustrated in Figure 4–8 exhibit diminishing returns to capital per worker but (by assumption) still have constant returns to scale. This latter characteristic means that if producers doubled all inputs (rather than just one input) then total output would also double.

Looking at the production function in Figure 4–8, we can see that at low levels of capital per worker (such as point *a*), new investment leads to relatively large increases in output per worker. But at higher levels of capital per worker (such as point *b*), the same amount of new investment leads to a smaller increment in output. Each addition of a unit of capital per worker (moving to the right along the *x*-axis) yields smaller and smaller increases in output per worker. More generally, giving the same number of workers more and more machinery yields smaller and smaller additions to output.

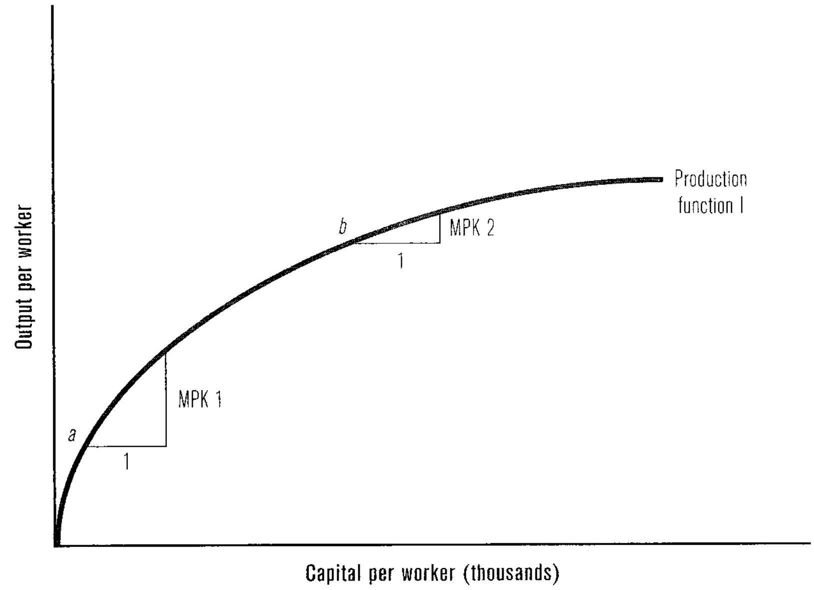

FIGURE 4-8 Diminishing Marginal Product of Capital
Along production function I, an addition of 1 unit of capital per worker at point *a* yields a much larger increase in output per worker than the same investment at point *b*. The incremental output produced by adding an additional unit of capital is called the marginal product of capital (MPK).

The assumption of a diminishing marginal product of capital has many implications, but three are particularly important for developing countries. Consider countries located toward the left on the *x*-axis of Figure 4-8, such as at point *a*. These countries have both relatively small amounts of capital per worker and low levels of output per worker. *The latter means, by definition, that these countries are relatively poor.* By contrast, countries toward the right have both higher levels of capital per worker and more output per worker, the latter implying that they are relatively rich. In general, low-income countries tend to have much less capital per worker than do richer countries. Therefore, if all else is equal between the two countries—a crucial qualifier—new investment in a poor country will tend to have a much larger impact on output than the same investment in a rich country. The three key implications are as follows:

- If all else is equal, poor countries have the *potential* to grow more rapidly than do rich countries. In Figure 4-8, a country located at point *a* has the potential to grow more rapidly than does a country at point *b* because the same investment will lead to a larger increase in output.
- As countries become richer (and capital stocks become larger), growth rates tend to slow. In other words, as a country moves along the production

function from point *a* to point *b* over a long period of time, its growth rate tends to decline.

- Because poor countries have the *potential* to grow faster than do rich countries, they can catch up and close the gap in relative income. To the extent this happens, income levels between rich and poor countries would converge over time.

These are very powerful implications. It is important to recognize that they rest on the assumption that all else is equal between the two countries, in particular that countries have the same rates of savings, population growth, and depreciation, and hence the same steady-state level of capital and income per capita. As this interpretation of the Solow model is not conditional on countries differing from one another in these key parameters, it is known as **unconditional convergence**. For all else to be equal, both countries also have to be operating along the same production function, have access to the same technology. If they are not, the predictions for rich and poor countries do not necessarily hold. For instance, a given poor country might actually be operating along different, and perhaps much flatter, production function than the one shown in Figure 4–8 if it does not have access to the same technology as reflected in the production function in Figure 4–8. In that case, each new investment (at a given level of the capital stock) would produce less output on the margin than that same additional investment would produce on the steeper production function. In that case, the poor country might not be expected to grow faster than the rich country and may never catch up. The phrase *ceteris paribus* (all else being equal), which is much used and often overlooked in economics, is of great importance in the convergence debate.

THE CONVERGENCE DEBATE

If it were true that poorer countries could grow fast while richer countries experience slower growth, poorer countries (at least those in which all else is equal) could begin to catch up and see their income levels begin to converge with the rich countries. Has this actually happened?

The short answer is that it has for some countries but not for most. Consider the example of Japan. In the 1960s, Japan's income per capita was only about 35 percent of average U.S. income, and it had a much smaller capital stock, giving it the potential for very rapid growth. (We use the United States as the benchmark for convergence because it has among the highest per capita incomes in the world and is usually considered the global technological leader.) Indeed, Japan's GDP growth rate exceeded 9 percent during the 1960s. By the time Japan had reached 70 percent of U.S. per capita income in the late 1970s, its GDP growth rate had slowed to about 4 percent. As its income continued to grow, its growth rate fell further, and growth was very slow after Japan reached about 85 percent of U.S. income in the early 1990s. Japan's experience illustrates the preceding three points very well: (1) When it was relatively poor, it could grow fast; (2) as its income increased, its growth rate declined; and (3) as a result, its income converged significantly toward U.S. income. People who boldly predicted in the 1960s and 1970s that Japan could grow at 7 to 9 percent per year indefinitely—and many people did—ignored the impact of diminishing returns of capital on long-term growth rates.

Japan is not the only country whose income has converged with the world leaders since 1960. Look again at the group of rapidly growing countries shown in Table 3-1. All these countries were relatively poor in 1960, and all grew by an average of between 3 and 6 percent per capita for nearly 50 years. Rich countries cannot grow that fast over a period of many years (in the absence of a continuous infusion of new technology), but poor countries can because they start with low levels of capital.

However, being poor and having low levels of capital per worker by no means guarantees rapid growth. As the upper sections of Table 3-1 show, many low-income countries recorded low growth. Not only did these countries not catch up but they fell further behind and their incomes diverged even more from the world leaders. The point is that low-income countries have the potential for rapid growth, *if* they can attract new investment and *if* that new investment actually pays off with a large increment in output.

Looking beyond the experience of a few individual countries, is there a general tendency for poor countries to grow faster and catch up with the richer countries? Broadly across all countries, the short answer is no. Figure 4-9 shows the initial level of per capita income in 1960 and subsequent rates of growth from 1960 to 2009 for

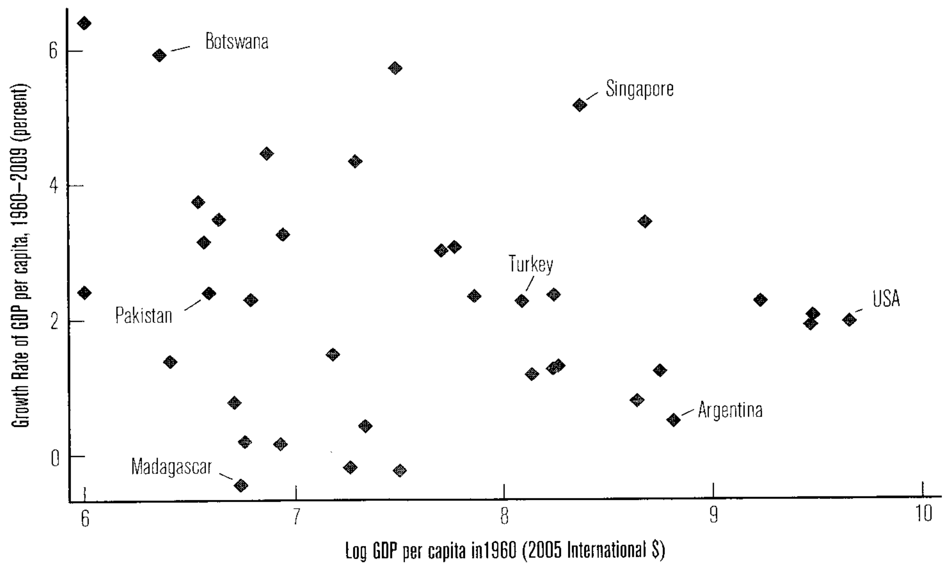

FIGURE 4-9 Gross Domestic Product (GDP) Growth, Unconditional

See Table 3-1 for data.

Source: Penn World Tables 7, http://pwt.econ.upenn.edu/php_site/pwt_index.php, accessed June 2011.

the 38 countries listed in Table 3-1. If it were true that poor countries were growing faster than rich countries, the graph would show a clear downward slope from left to right. Poor countries would record a high rate of growth (and appear in the upper left part of the figure) and rich countries would display a slower growth rate (and be in the bottom right). But there is no clear pattern evident in the figure. For example, income per capita in 1960 was approximately the same in Botswana, Pakistan, and Madagascar; yet while Botswana subsequently grew at nearly 6 percent per year, Pakistan's subsequent growth was only 2.3 percent per year and Madagascar's income per capita actually declined. Similarly, Pakistan, Turkey, and the United States all grew at roughly the same rate from 1960 to 2009, yet they began the period with quite different levels of income. The only part that seems accurate is that almost all the rich countries display relatively slow growth rates, as expected. Such results have been documented in many studies using larger samples and more sophisticated statistical techniques. The empirical fact is clear: There has been no *general* tendency for poor countries to catch up to the world leaders. If anything, the opposite has been true. As we saw in Figure 2-1, for the last two centuries, the gap between the richest and the poorest regions of the world has grown, implying a divergence of incomes for these countries.

However, the simple graph in Figure 4-9 does not really do justice to the predictions of convergence, which are based on the critical assumption that all else is equal across countries. This assumption clearly is not true for all countries in the world. Instead, if convergence were to occur, we should expect to find it among countries that share some broad key characteristics, such as a similar underlying production function and similar rates of saving, population growth, depreciation, and technology growth. Some of the poor countries in Figure 4-9, for example, have low saving rates or very little growth in technology compared with other countries and, therefore, have much less potential for rapid growth. To see if the convergence predictions of the Solow model hold under these stricter conditions, we have to dig a little deeper.

Figure 4-9 demonstrates the lack of unconditional convergence, *unconditional* in the sense that it imposes the assumption that all countries share the same key parameters (depreciation, savings, and population growth rates in particular). Digging deeper into the Solow model's predictions, however, we consider the implications for convergence of allowing each country to take on its own individual combination of these key parameters. This allows each country to have its own individual steady-state level of capital and income per capita. Once we allow that, the question is whether initially poorer countries grow faster as they approach their own steady state. Figure 4-10 applies statistical techniques to relax the previous assumption that all countries have the same rates of population growth. Looking at the data from the same countries as in Figure 4-9, but conditional on the countries' population growth rates, we begin to see the predicted negative association between initial income levels and subsequent growth rates. This is known as **conditional convergence**.

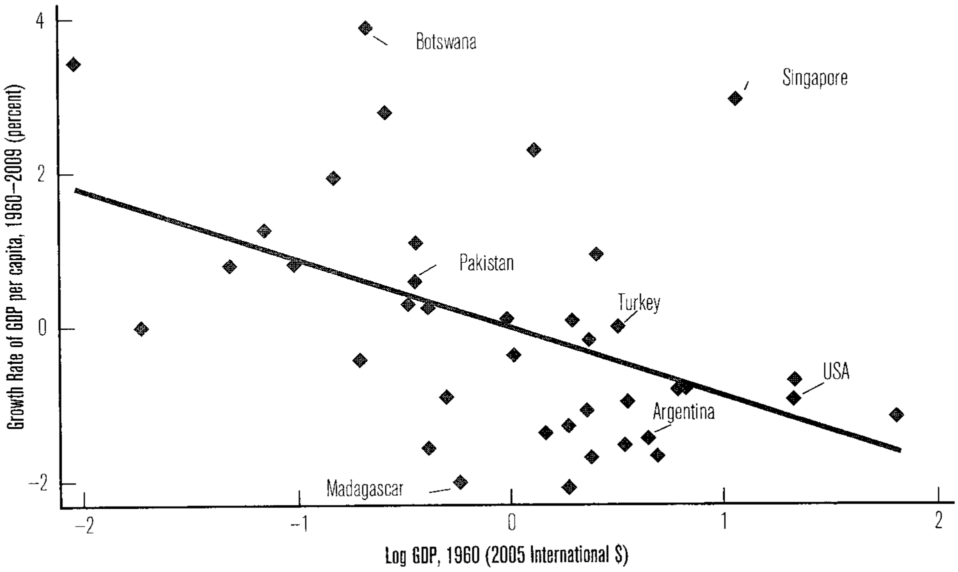

FIGURE 4-10 Gross Domestic Product (GDP) Growth, Conditional on Population Growth

See Table 3-1 for data.

Sources: Penn World Tables 7, http://pwt.econ.upenn.edu/php_site/pwt_index.php, accessed June 2011; World Bank, "World Development Indicators," http://databank.worldbank.org.

Another approach is to pick a group of countries from all income levels that are similar in their policy choices, geographic characteristics, or some other variable. Economists Jeffrey Sachs and Andrew Warner, for example, examined the evidence for convergence among all countries that had been consistently open to world trade since 1965.[21] Open economies, as discussed in Chapter 3, are similar in that they have similar (global) markets for their products, purchase their inputs on world markets, and can acquire new technology relatively quickly from other open economies through imports of new machinery and their connections to global production networks. What if we include countries' openness to trade among the conditioning variables? Figure 4-11 conditions GDP growth on countries' rates of savings and population growth in addition to the proportion of time between 1960 and 2009 that they were open to trade. Clearly, these additional conditioning variables sharpen the inverse correlation between initial income and subsequent growth as indicated by the tighter fit around the trend line.[22] Initially wealthier countries may still grow faster in absolute terms than initially poorer countries if the former are converging toward

[21] Jeffrey D. Sachs and Andrew Warner, "Economic Reform and the Process of Global Integration," *Brookings Papers on Economic Activity*, 26, no. 1 (1995), 1–118.

[22] Tracing the countries highlighted in Figure 4-9 through Figures 4-10 and 4-11, we note that, despite the additional conditioning variables, Botswana and Singapore consistently grew faster than predicted for their initial incomes (indicated by their positions above the trend lines), whereas Madagascar and Argentina consistently grew more slowly than predicted.

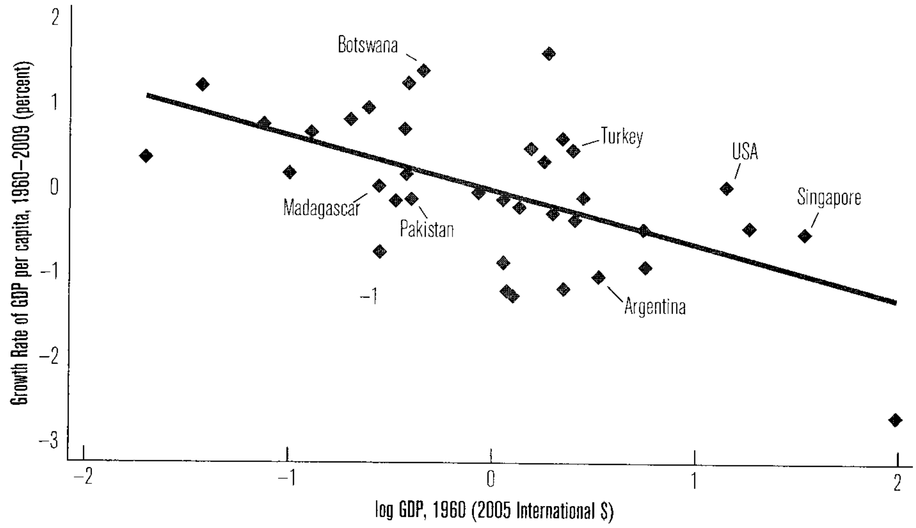

FIGURE 4-11 Gross Domestic Product (GDP) Growth, Conditional on Openness, Savings, and Population Growth

See Table 3-1 for data.

Sources: Penn World Tables 7, http://pwt.econ.upenn.edu/php_site/pwt_index.php, accessed June 2011; World Bank, "World Development Indicators," http://databank.worldbank.org.

much higher steady-state levels of capital per capita. That is why we do not observe unconditional convergence across rich and poor countries' levels of income. On the other hand, if each country is allowed to have its own steady-state level, we do see faster growth among initially poorer countries, as predicted by the Solow model and its implication of conditional convergence.

BEYOND SOLOW: NEW APPROACHES TO GROWTH

A new generation of models takes off where Solow left off, by moving beyond the assumptions of an exogenously fixed saving rate, growth rate of the labor supply, workforce skill level, and pace of technological change. In reality, the values of these parameters are not just given but are determined partially by government policies, economic structure, and the pace of growth itself. Economists have begun to develop more-sophisticated models in which one or more of these variables is determined within the model (that is, these variables become endogenous to the model).[23]

[23]The seminal contributions to the new growth theory are Paul Romer, "Increasing Returns and Long-Run Growth," *Journal of Political Economy* 94 (October 1986), 1002–37; Robert Lucas, "On the Mechanics of Economic Development," *Journal of Monetary Economics* 22 (January 1988), 3–42; and Paul Romer, "Endogenous Technological Change," *Journal of Political Economy* 98 (October 1990), S71–S102.

These models depart from the Solow framework by assuming that the national economy is subject to **increasing returns to scale**, rather than constant returns to scale. A doubling of capital, labor, and other factors of production leads to *more* than a doubling of output. To the extent this occurs, the impact of investment on both physical capital and human capital would be larger than suggested by Solow.

How can a doubling of capital and labor lead to more than a doubling of output? Consider investments in research or education that not only have a positive effect on the firm or the individual making the investment but also have a positive "spillover" effect on others in the economy. This beneficial effect on others, called a **positive externality**, results in a larger impact from the investment on the entire economy. The benefits from Henry Ford's development of the production line system, for example, were certainly large for the Ford Motor Company, but they were even larger for the economy as a whole because knowledge of this new technique soon spilled over to other firms that could benefit from Ford's new approach.

Similarly, investments in research and development (R&D) lead to new knowledge that accrues not only to those that make the investment but to others who eventually gain access to the knowledge. The gain from education is determined not just by how much a scientist's or manager's productivity is raised by investment in his or her own education. If many scientists and managers invest in their own education, there then will be many educated people who will learn from each other, increasing the benefits from education. An isolated scientist working alone is not as productive as one who can interact with dozens of well-educated colleagues. This interaction constitutes the externality. In the context of the "sources of growth" analysis introduced in Chapter 3, such externalities suggest that the measured contribution of physical and human capital to growth may be larger than that captured by the Solow framework. Among other implications, this outcome could account for a significant portion of the residual in the Solow accounting framework, meaning that actual TFP growth is smaller than many studies have suggested.

Another important implication is that economies with increasing returns to scale do not necessarily reach a steady-state level of income as in the Solow framework. When the externalities from new investment are large, diminishing returns to capital do not necessarily set in, so growth rates do not slow, and the economy does not necessarily reach a steady state. An increase in the saving rate can lead to a *permanent* increase in the rate of economic growth. These models can explain continued per capita growth in many countries without relying on exogenous technological change. Moreover, they do not necessarily lead to the conclusion that poor countries will grow faster than rich countries because growth does not necessarily slow as incomes rise, so there is no expectation of convergence of incomes. Initial disparities in income can remain, or even enlarge, if richer countries make investments that encompass larger externalities.

Because growth can perpetuate in these models without relying on an assumption of exogenous technological change, they often are referred to as **endogenous**

growth models. They are potentially important for explaining continued growth in industrialized countries that never reach a steady state, especially those engaged in R&D of new ideas on the cutting edge of technology.

For developing countries, the new models reinforce some of the main messages of the Solow and Harrod-Domar models. Like their forerunners, these models show the importance of factor accumulation and increases in productivity in the growth process. The potential benefits from both of these sources of growth are even greater in endogenous growth models because of potential positive externalities. The core messages of saving, investing in health and education, using the factors of production as productively and efficiently as possible, and seeking out appropriate new technologies are consistent across all these models.

The applicability of endogenous growth models to developing countries is questionable, however, because many low-income countries can achieve rapid growth by adapting the technologies developed in countries with more advanced research capacities rather than making the investments in R&D themselves. For many low-income countries, the Solow model's assumptions of exogenous technological change and constant returns to scale in the aggregate production function may be more appropriate. Productivity growth in such countries may also depend as much on appropriate complementary investments and policies as on technological change itself.

SUMMARY

- Formal growth models provide a more precise mechanism to explore the contributions to economic growth of both factor accumulation and productivity gains. These models allow us to understand better the implications of changes in saving rates, population growth rates, technological change, and other related factors on output and growth.
- The Harrod-Domar model assumes a fixed-coefficients production function, which helps simplify the model but introduces strict rigidity in the mix of capital and labor needed for any level of output. In this model, growth is directly related to saving in inverse proportion to the incremental capital-output ratio.
- The Harrod-Domar model usefully emphasizes the role of saving, but at the same time overemphasizes its importance by implying that saving (and investment) is sufficient for sustained growth, which it is not. Also, the model does not directly address changes in productivity. In addition, the model's assumption of a fixed ICOR leads to increasing inaccuracy over time as the structure of production evolves and the marginal product of capital changes.

- The Solow model improves on some of the weaknesses in the Harrod-Domar framework and has become the most influential growth model in economics. The model allows for more flexibility in the mix of capital and labor in the production process and introduces the powerful concept of diminishing marginal product of capital. It allows for exploration of the impact on growth of changes in the saving rate, the population growth rate, and technological change. The model helps provide a deeper understanding of a much wider range of growth experiences than the Harrod-Domar model. Nevertheless, the model does not provide a full explanation for growth. It does not provide insights into the more fundamental causes of factor accumulation and productivity growth and, as a one-sector model, does not address the issue of resource allocation across sectors.
- The Solow model has several powerful implications, including (1) poor countries have the *potential* to grow relatively rapidly; (2) growth rates tend to slow as incomes rise; (3) across countries that share important common characteristics, the incomes of poor countries potentially can converge with those of the rich countries; and (4) acquiring new technology is central to both accelerating and sustaining economic growth.
- Some poor countries have achieved rapid growth and seen their incomes converge with those of richer countries, but many have not and have seen their incomes fall further behind. There is no evidence of unconditional convergence across countries, but there is strong evidence of *conditional* convergence, in which countries sharing certain characteristics are able to achieve rapid growth and begin to catch up with the richer countries. We find evidence of conditional convergence in the inverse relationship between initial income and subsequent growth rates, when we account for countries differing in key parameters and hence in their steady-state levels.
- In the 1990s, new theories of economic growth gained popularity. These theories sought to fill in key gaps in the neoclassical model. In particular, these models sought to explain the rate of technological progress, which Solow had identified as the determinant of long-term growth in per capita income but that remained exogenous in his model. These new growth theories are thus known as endogenous growth models.