development activities, or unemployment), and how the quantities of inputs available in the economy and the economy's technological options evolve over time. We call these behavioral assumptions, because they describe how the economy's firms choose what to produce, how to produce it, and what investments to make; how the economy's households choose where to work, what to consume, and what to save; and how markets work (or fail to work) when producers and consumers interact with each other. The simplest models assume that firms and households follow simple rules of thumb, such as the rule that households always save a certain percentage of their income, and assume that key features of the economy (such as the level of technology) always grow at exogenously fixed rates. More elaborate models build up assumptions from more detailed and thorough-going microeconomic foundations, describing producers and households as seeking to maximize profits or utility subject to the constraints they face.

Behavioral assumptions, too, have evolved over time in several important ways. First, the earliest models assumed that producers and households had little ability or inclination to modify their choices in response to economic conditions and that as a result, markets would fail to guide all the economy's inputs into their most productive uses. These were replaced by models making almost diametrically opposite assumptions, in which producers and households were assumed to face great flexibility and markets were assumed to function perfectly. More recently yet, these have been replaced by models staking out a more nuanced, intermediate position, in which producers and households are flexible and responsive to economic conditions but in which markets nonetheless might fail to direct all inputs into their best possible uses.

Second, the earliest models largely took technological conditions as fixed and unchanging. These were replaced by models in which technology was assumed to improve over time, for reasons that were unexplained within the model. More recently, model builders have tried to identify actors within their models who are responsible for generating technological improvements and to formulate assumptions about how these actors make the decisions that determine the rate of **technical change**.

Model builders make their assumptions specific and clear by writing them down as systems of mathematical equations. These equations involve **exogenous parameters**, whose values are taken as given from outside the model, and **endogenous variables**, whose values are determined (as functions of the exogenous parameters) by solving the system of equations. The most important endogenous variable in all these models is, of course, the rate of economic growth. Early models treated only the growth rate as endogenous, but later models have treated a growing array of variables—such as rates of investment and technical change—as jointly endogenous with the growth rate.

For each model we examine below, we describe the technological and behavioral assumptions, and their implications for how the growth rate would evolve over time if the parameters of the model remained constant and how a country's growth trajectory would change as a result of changes in key exogenous parameters. We embed this technical discussion of specific models into a larger description of the intellectual interplay between theoretical predictions and empirical observations that has driven the evolution of thought about economic growth.

4.2 The 1940s and 1950s: Capital Fundamentalism, Structuralism, and Dualism

4.2A Themes

Economic thought regarding growth in developing countries in the 1940s and 1950s was characterized by capital fundamentalism, structuralism, and dualism. **Capital fundamentalism** refers to the belief that the most important proximate source of economic growth is the accumulation of physical capital. The emphasis economists placed on capital accumulation in the

1940s is understandable, given the perception at the time that the Soviet Union's recent industrialization through rapid (and forced) saving and physical capital investment had been a great success.

Structuralism refers to the belief that the world is a rigid place, in which economic actors—especially the poor and tradition-bound peasants who were thought to populate developing countries—have little ability or inclination to alter what they do in response to economic incentives. When, for example, labor is plentiful and capital is scarce, the price of capital might rise, but such price increases would not set into motion self-equilibrating market processes, in which people decide to invest in the creation of more capital or to shift to production methods that make greater use of cheap labor and rely less on costly capital. In such a world, unemployment can remain rampant for long periods without markets showing any tendency to absorb the unemployed. This view was rendered compelling by the recent experience of the Great Depression.

Dualism refers to the belief that developing economies must be understood as having two sectors, traditional and modern, each with peculiar characteristics and very different growth prospects. The large traditional sector, often identified with agriculture, was assumed to be "backward," in the sense that production was overseen by poor, irrational peasants who were not driven by modern profit-maximizing motivations, who operated at very low levels of productivity, and who would have little prospect for improving their productivity even if they were inclined to try. The modern sector, often identified with manufacturing, was by contrast seen as a potentially dynamic sector, though still small in most developing countries. It was in modern manufacturing that physical capital accumulation was thought to hold tremendous promise of increasing productivity and incomes, and where producers were more inclined to plow growing profits back into saving and investment.

The predominant perspective on economic growth during this period had profound policy implications.

In the 1940s and 1950s, capital fundamentalism motivated emphasis on raising rates of saving and physical capital accumulation. Structuralism motivated extensive government involvement in bringing those increases about. Dualism dictated that the new capital be concentrated in the modern manufacturing sector and that such "industrialization" be encouraged through subsidies and by protecting domestic manufacturing enterprises from international competition.

The following sections describe two influential theoretical growth models of the 1940s and 1950s. The Harrod–Domar model, which influenced thinking about growth in developed as well as developing countries, exhibits capital fundamentalist and structuralist thinking. The Lewis model incorporates dualism as well and was considered useful primarily for understanding developing countries.

4.2B The Harrod–Domar growth model

The ideas underlying what is now known as the Harrod–Domar growth model were introduced independently in papers by Roy Harrod (1939) and Evsey Domar (1946). At the center of the Harrod–Domar model is a very simple and extreme assumption regarding the economy's production technology: A fixed quantity u of labor and a fixed quantity v of physical capital are required to produce each unit of GDP. That is, the aggregate production function is assumed to take the following extreme form:

$$Y = min\left[\left(\frac{1}{u}\right)L, \left(\frac{1}{v}\right)K\right] \tag{4.1}$$

where Y is total GDP, L is the total quantity of labor, and K is the total quantity of physical capital. The *min*[.,.] function indicates that total output is equal to the smaller of the two arguments. The parameters u and v are known as the *labor-output ratio* and the *capital-output ratio* and are fixed technological parameters defining an inflexible technology.

Notice several things about the technological assumptions embodied in equation 4.1. First, the production function assumes that capital and labor are the only important inputs to production. Second, it assumes that technological options are few and rigid. Indeed, labor and capital must always be combined in exactly the same proportion (u/v units of labor per unit of capital) to produce a unit of output; for this reason the technology is labeled a **fixed proportions production technology**. If the ratio of labor to capital in the economy is greater than u/v, then some labor has no use in production and will remain unemployed, and if the ratio of labor to capital is less than u/v, some capital will remain unemployed. Third, it assumes that production is characterized by constant returns to scale.

The model's behavioral assumptions offer simple descriptions of how K and L evolve over time and how given stocks of K and L are utilized at any point in time. The capital stock K is increased through investment. In an economy that is closed to international investment, increases in K must be financed by local saving. The critical behavioral assumption driving growth in K, then, is the assumption that people save a fixed proportion s of their total income. Capital accumulates when this saving is greater than the rate at which old capital depreciates. The labor force L is assumed to grow at an exogenously fixed rate n.[1] Significantly, the technological parameters u and v are taken as exogenously fixed, with no built-in tendency to evolve over time.

Labor is assumed (at most times) to be abundant relative to capital, in the sense that the current stock of capital is too small to provide productive employment for all workers. (This notion of labor abundance makes sense only as a result of the fixed proportions production technology, which implies that the current stock of capital K can provide productive employment for only $(u/v)K$ units of labor.) In practice, then, output can be treated as a function of K alone, because the economy's stock of labor never represents a constraint on production.

The assumptions of the previous several paragraphs may be summarized by a simple set of equations. Combining the technological assumptions with the assumption that labor never represents a constraint on production, we can state the economy's effective aggregate production function as

$$Y = \left(\frac{1}{v}\right)K \tag{4.2}$$

This says that total GDP is just a constant multiple of the size of the capital stock. Denoting the derivative of a variable x with respect to time by $\dot{x}$, we can restate equation 4.2 in a way that illustrates the link between *increases* in Y and *increases* in K:

$$\dot{Y} = \left(\frac{1}{v}\right)\dot{K} \tag{4.3}$$

That is, Y rises and falls strictly in proportion to K. Notice that as long as labor remains abundant relative to capital (so that this equation provides a good description of growth in aggregate production), the economy is characterized by constant marginal returns to capital: Each additional unit of capital increases Y by the same amount.

In light of equation 4.3, if we are to understand growth in Y, we must understand growth in K. Increases in the capital stock ($\dot{K}$) are achieved when the rate of investment exceeds the rate of

[1] Throughout this chapter we essentially treat the labor force as a fixed proportion of the total population, in which case the growth rate of the labor force is just the population growth rate, and growth in output per person is identical to growth in output per worker.

depreciation of the old capital stock. Invoking the assumption that people always save a fixed proportion s of their income and letting d represent the fraction of the old capital stock that decays or becomes obsolete each year, we can describe the change in size of the capital stock in a short interval of time as

$$\dot{K} = sY - dK \tag{4.4}$$

Equations 4.3 and 4.4 fully describe the assumptions of the model. We solve the model, seeking to understand its implications for growth, by solving this set of equations for a statement describing how the rate of economic growth ($\dot{Y}/Y$) is determined as a function of the key exogenous parameters of the model (v, s, and d). Substituting equation 4.4 into equation 4.3, we find that

$$\dot{Y} = \left(\frac{1}{\nu}\right)(sY - dK) \tag{4.5}$$

or (dividing both sides by Y, and recognizing that $\nu = K/Y$)

$$\frac{\dot{Y}}{Y} = \frac{s}{\nu} - \left(\frac{d}{\nu}\right)\left(\frac{K}{Y}\right) = \frac{s}{\nu} - d \tag{4.6}$$

Equation 4.6 describes how the model's main endogenous variable, the growth rate of total GDP, is determined as a function of the model's parameters. (Later models focus on the growth rate in per capita rather than total GDP.) We conclude that as long as the parameters remain fixed, output growth proceeds at a constant rate. Differences in growth rates across countries must be attributed to differences in the saving rate s or the capital-output ratio v, and increasing a country's growth rate would require increasing s or reducing v (assuming the rate of depreciation is unalterable).

How might changes in s or v come about? In the model's world of rigidities and poorly functioning markets, increases in s are unlikely in the absence of intervention by developing country governments, and these governments might require assistance (i.e., foreign aid) from the world's rich countries. Reductions in the technological parameter v, which require improvements in efficiency or technology, were thought even more difficult to engineer even by government. The main conclusion of the model thus comes as little surprise given the assumptions on which it is built: To increase the GDP growth rate, a country's government must increase the rate of investment by the public sector, financing that investment through taxes and foreign aid.

The model has additional implications that are less obvious and that are less compelling in light of historical experience. In this model, the growth of K (and thus the rate of growth in the number of productive jobs) is governed by the saving rate, while growth in L (the number of workers) is governed by unrelated forces. It can be shown (see problem 1) that K (like Y) grows at the rate $(s/\nu) - d$, and the labor force grows at the rate n, which need not equal $(s/\nu) - d$. If K grows more rapidly than L then the number of jobs in the economy grows more rapidly than the number of workers, and unemployment falls. Under such circumstances, GDP per capita rises, even though the productivity of each employed worker remains fixed, simply because the share of workers who are in unproductive unemployment falls. If capital continues to grow more rapidly than the labor force, production might eventually use up all available labor. In this case, further capital accumulation bids up wages and prices, generating inflation without increasing output per capita, and growth in GDP per capita ceases. If, however, K grows less rapidly than L, then unemployment must rise and GDP per capita falls. Unless $s/\nu - d$ just happens to equal n, the economy will be marked by rising or falling unemployment. This **knife-edge** property of the model implies perpetual instability. Furthermore, if $s/\nu - d$ did just equal n, total GDP would rise but GDP per capita would remain constant. As we will see, the mismatch between these implications and real-world experience contributed to the eventual rejection of the Harrod–Domar model.

Such wage-growth patterns fall far short of confirming the Lewis model's assumptions or policy implications, because similar patterns of wage growth could be generated by very different models. For example, wage growth could follow this pattern even in models in which all sectors are capitalist and all workers and producers have equal propensities to save, as long as incipient wage increases lead to large increases in household labor supply at low levels of income and smaller increases in the hours of labor supplied at higher income levels. In such models, initial growth-induced increases in the demand for labor would lead to rapid increases in labor hours and slow increases in wages, whereas later growth and labor-demand expansion would induce more rapid wage growth and less expansion of labor hours. The models would thus explain wage growth patterns just as well as the Lewis model but would offer no justification for subsidizing modern sector expansion and would yield no prediction that expansion of a modern sector would raise rates of saving, investment, and growth.

Though we now reject the extreme assumptions of the Lewis model and many of its policy implications, contemporary development thinking continues to be shaped by Lewis's accurate observations that growth is almost always accompanied by structural change, that the elasticity of labor supply helps determine the extent to which the benefits of growth are shared by workers throughout the economy, and that wages sometimes tend to differ among identical workers in different sectors or geographic locations. All of these point to the need for careful study of labor markets in developing countries, as in Chapter 9.

4.3 The 1960s and 1970s: Neoclassical Perspectives, Technology, and Human Capital

4.3A Themes

Research in the 1960s and 1970s chipped away at the theoretical and empirical underpinnings of capital fundamentalism, structuralism, and dualism. Capital fundamentalism gave way to interest in the combined contributions of physical capital accumulation, technical change, and human capital accumulation. The introduction of a new theoretical growth model, the neoclassical growth model, raised theoretical questions about the potential for physical capital accumulation alone to explain the sustained growth observed in the United States and other countries and pointed toward technical change as an additional source of growth. Simultaneous developments in the study of labor markets and wage formation pointed to the importance of human capital accumulation. Growth accounting exercises (as discussed in Chapter 3) provided further reason to believe that human capital accumulation and technical change were at least as important as physical capital accumulation in explaining growth.

Structuralist views of developing country markets diminished as collection and analysis of data from farms and households in developing countries demonstrated that even poor rural families respond to economic incentives in ways predicted by models of rational economic behavior. The structuralist assumption that poorly functioning markets should be overridden by extensive government intervention was also called into question by highly disappointing experiences with government planning. Indeed, the intellectual pendulum swung to the opposite extreme in this period, as a growing number of development analysts argued for withdrawal of government intervention and greater reliance on markets for allocating resources and motivating growth.

Dualism of the sort exhibited in the Lewis model lost favor in this period, too, for a variety of reasons. Research suggested that the marginal product of labor is not zero in agriculture, that agriculture can be a dynamic source of saving and physical capital accumulation, and that investment in physical capital can induce productivity increases in agriculture as well as manufacturing. The experience of the Green Revolution (see Chapter 20), moreover, demonstrated the great potential for new technologies to increase agricultural productivity. On top of this, it was becoming apparent that many of the poor in developing countries earned their livelihoods in agriculture and would not begin to share in the benefits of growth and development for a very long time unless new efforts were made to increase agricultural productivity and income. For a more detailed discussion of changing economic thought regarding growth and development in this period, see Little (1982).

The policy implications of the new perspective on economic growth were very different from those of earlier years.

> In the 1960s and 1970s, new views of economic growth motivated greater interest in encouraging human capital investment and technical change, greater attention to agriculture, and reduced government involvement in the economy.

In what follows, we discuss three versions of the highly influential neoclassical growth model. The first is the "simple" model without technical change (or human capital accumulation) that Robert Solow used to demonstrate the inadequacy of physical capital accumulation alone to explain the real-world experience of sustained economic growth. The second is the model *with* technical change, in which Solow showed that introducing technical change solves the puzzle presented by the simpler model. The final version acknowledges that the accumulation of human as well as physical capital might be important in economic growth.

4.3B The simple neoclassical growth model

In his landmark paper "A Contribution to the Theory of Economic Growth" (1956), Robert Solow introduced what we now know as the neoclassical growth model, which served as the unifying framework for thought about economic growth for several decades. He motivated the new model by pointing out two reasons for dissatisfaction with the then-popular Harrod–Domar growth model. First, that model rests on stark technological assumptions that contemporary economists found implausible. Second, the knife-edge property of that model implies a perpetual macroeconomic instability that seemed counter to observed experience, at least in the United States.

Solow pointed out that the rigid assumptions and the unsatisfying implications of the Harrod–Domar model are closely interrelated. By assuming a production technology in which labor and capital can only be productively employed in exactly fixed proportions, the model assumes away any potential for producers to respond to capital scarcity by using more labor-intensive methods of production. If some workers are unemployed, we might expect them to offer to work at lower wages, rendering labor cheaper relative to capital. Producers who can choose from multiple production methods would be encouraged by the lower wages to switch to production methods that require more labor and less capital per unit of output. By so doing, they would increase the number of workers who can find employment with the fixed capital stock. If markets work perfectly, this process should continue until unemployment disappears. Such equilibrating changes were ruled out by the Harrod–Domar model's rigid technological assumptions.

Solow decided to examine what would happen to the predictions of the Harrod–Domar model as a result of just two changes in its assumption (one technological and one behavioral). First, he replaced the rigid fixed-proportions production technology by a **variable proportions** or **neoclassical production function**, in which output can be produced with varying combinations of labor and capital and each input is subject to diminishing marginal returns (assumptions that had long been employed in much economic analysis outside the areas of growth and development), Second, he replaced the assumption of extensive unemployed labor by the assumption that rational producers respond to signals in perfectly functioning markets in ways that drive the economy to full employment of both labor and capital at all times. Solow continued to assume, along with Harrod and Domar, that technology exhibits constant returns to scale, that saving is a fixed proportion of income, and that the labor force grows at an exogenously determined rate n.

Solow showed that introducing such changes to the model's assumptions eliminated its undesirable predictions of instability. Unemployment is eliminated from the model, and capital can accumulate relative to labor with no potential for a sudden deceleration of growth when unemployed labor is used up.

Solow also demonstrated, however, that making these changes leads to important changes in some of the model's predictions that had not previously been questioned. In the Harrod–Domar model, a constant saving rate can serve as the source of steady, sustained growth (as long as the economy doesn't run out of unemployed labor), and increases in the saving rate lead to sustained increases in the growth rate. We will see below that in Solow's neoclassical model, a constant rate of saving by itself cannot sustain growth in the long term, and permanent increases in the saving rate produce only temporary boosts in the rate of growth.

The source of this dramatic change in theoretical prediction is the assumption of neoclassical production technology, characterized by the diminishing marginal returns to capital. In the presence of such diminishing returns, a steady rate of saving implies steady additions to the capital stock as a proportion of output, but this translates into a *declining rate* of increase in capital per worker and a declining rate of economic growth, because as the quantity of capital per worker increases, diminishing returns dictate that additional units of capital lead to smaller and smaller increases in output per worker. In fact, growth in output per worker based entirely on saving and capital accumulation is bound eventually to grind to a halt in the simple neoclassical model. Increases in savings rates might boost growth in the short run and allow the economy to achieve a higher *level* of per capita GDP before economic growth halts, but such growth would ultimately cease. We demonstrate this formally below.

Solow did not stop there, however. He pointed out that the revised model, too, failed to match historical experience. The United States, for example, *had* experienced a nearly constant saving rate *and* steady growth over a very long period. He concluded that if growth based on steady capital accumulation alone is bound to diminish over time, then U.S. growth must have been based on something more than just steady capital accumulation. The something more to which he drew attention was technical change.

In this section, we describe what happens when we introduce Solow's neoclassical modifications of the Harrod–Domar model without introducing the possibility of technical change. We will examine this simple model's implications, including the puzzling prediction that growth must eventually grind to a halt.

We begin with the assumptions that the economy's aggregate production function is neoclassical and that the economy's stocks of labor and capital are fully employed at all times. These assumptions together underlie the assumption that

$$Y = F(K, L) \tag{4.7}$$

where Y is GDP and K and L are the economy's stocks of capital and labor inputs. The production function $F(.)$ is assumed to be characterized by diminishing marginal returns to either input. The production function furthermore exhibits constant returns to scale. The economy's many owners of capital and labor resources are assumed to sell capital and labor services to producers in perfect factor markets. In perfect markets, labor and capital are fully employed and are allocated across firms in ways that maximize their productivity. All this happens regardless of how ownership of the factors is distributed, and thus total production in the economy depends only on the total quantities of K and L and not on how their ownership is distributed. The rest of the model is the same as in Harrod–Domar: L is assumed to grow at a fixed rate n, K grows through capital accumulation financed by saving (in excess of the rate of depreciation d), and saving is a constant proportion s of income.

Notice that the only way for output per worker ($y = Y/L$) to grow in this model is through increases in the capital-labor ratio ($k = K/L$), which indicates the quantity of capital available per worker in the economy. Given our interest in the growth of per capita GDP, it will be useful to recast our technological assumptions in per-worker terms. Dividing both sides of equation 4.7 by L and invoking the assumption of constant returns to scale, we find that

$$y = \frac{Y}{L} = \left(\frac{1}{L}\right)F(K, L) = F\left(\frac{K}{L}, 1\right) = f(k) \tag{4.8}$$

where the last equality simply assigns a new name to the function describing the relationship between capital per worker and output per worker. This new function $f(.)$ relates the average product of labor y to the level of capital per worker k. As illustrated in Figure 3.3, y must increase with k, but it must increase at a decreasing rate as a result of diminishing marginal returns to capital.

Given the close link between the capital-labor ratio k and per capita production y, we can understand the model's implications for the rate of economic growth by understanding what governs the rate of change in k, which we denote by $\dot{k}/k$. It may be shown that the rate of growth of the capital-labor ratio is just the rate of growth of capital minus the rate of growth of labor, as in

$$\frac{\dot{k}}{k} = \frac{\dot{K}}{K} - \frac{\dot{L}}{L} \tag{4.9}$$

As in the Harrod–Domar model, $\dot{K}$ is equal to $sY - dK$, where Y now equals $F(K, L)$ and the rate of growth of the labor force is n. After making these substitutions, equation 4.9 becomes

$$\frac{\dot{k}}{k} = \frac{sF(K, L)}{K} - d - n \tag{4.10}$$

Noting that $F(K, L)$ may be written as $LF(k, 1) = Lf(k)$ (making use of the assumption of constant returns to scale), and multiplying both sides of the equation by k, we find the equation embodying the core implications for growth of the simple neoclassical model:

$$\dot{k} = sf(k) - k(n + d) \tag{4.11}$$

The left side is the change in k experienced in any period. If this is positive, k is rising; if it is negative, k is falling. The right side relates this change to current economic circumstances. The first term on the right side describes the saving and investment forthcoming per person at any current level of k. The second term indicates the investment per person required simply to maintain the entire labor force at the current level of k, given the need to replace depreciating capital and the need to equip additional workers entering the market through labor force growth. The equation tells us that if per-worker investment forthcoming is greater than the per-worker investment required to maintain a constant level of capital per worker, then the capital-labor ratio (and output per worker) will rise. If the per-worker investment forthcoming is less than $k(n + d)$, however, then the capital-labor ratio will fall.

One aim in analyzing a dynamic model like this is to determine the rates at which key variables (such as k and y) would grow in long-run **steady-state equilibrium**, which is achieved when the economy evolves to a point at which growth rates cease to change. To identify steady-state equilibrium in this model, it is useful to employ the diagram found in Figure 4.2a. The horizontal axis measures the capital-labor ratio k. The vertical axis is measured in units of output per worker, which are also the units in which we measure capital, saving, and investment. As indicated by the $f(k)$ graph, output per worker rises at a decreasing rate as k increases, because capital's marginal product is positive but diminishing. The figure describes what happens to the two components of the right side of equation 4.11 as k increases. Because $sf(k)$ is just a constant multiple (between zero and one) of $f(k)$, diminishing marginal returns to capital imply that the $sf(k)$ graph, too, has positive but decreasing slope. Since n and d are constants, $k(n + d)$ is a straight line out of the origin with slope $n + d$.

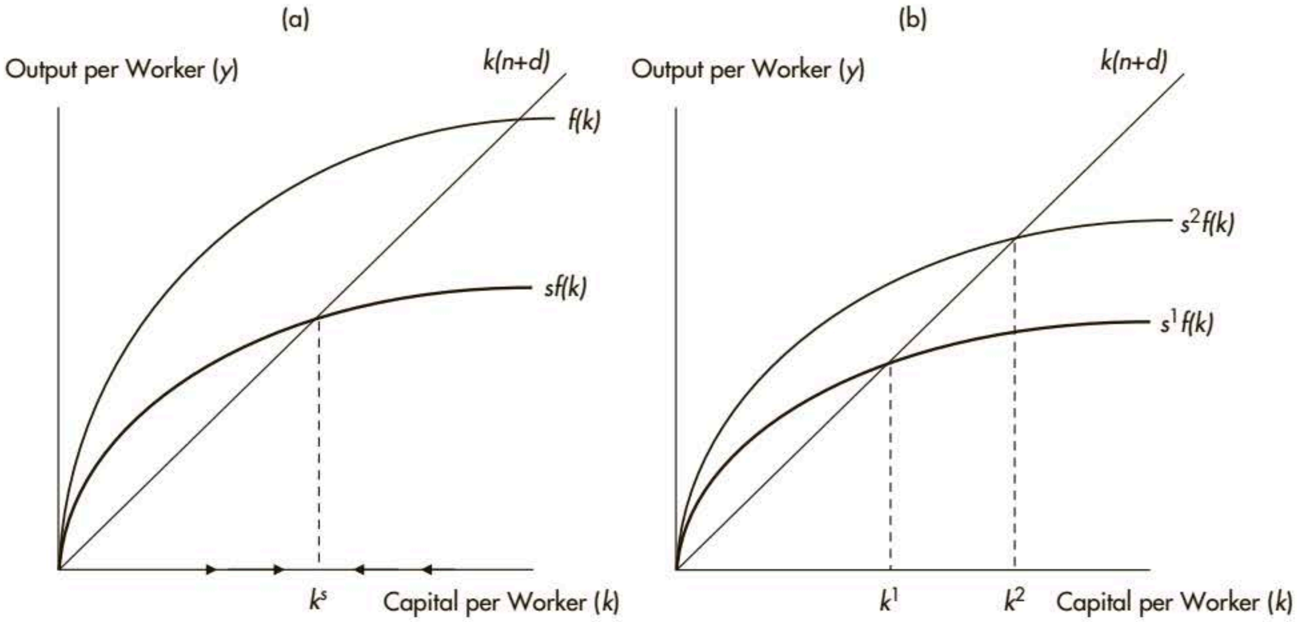

FIGURE 4.2
Steady-State Equilibrium in the Simple Neoclassical Growth Model

We can use Figure 4.2a to demonstrate a tendency for the economy to progress toward a steady-state equilibrium in which the capital-labor ratio and output per worker are unchanging.[2] That is, the economy will progress toward an equilibrium in which there is no (per capita) economic growth! To see this, notice first that if the economy ever reached the level of capital per worker k^s associated with the intersection of the $sf(k)$ and $k(n+d)$ graphs, the economy would remain at that level of k forever. At that level, the amount of investment forthcoming in the economy, $sf(k^s)$, is just enough $(k^s(n+d))$ to replace depreciated capital and to provide new workers with the same level of capital per person as the old workers already have. According to equation 4.11, $\dot{k}$ would be zero, and there would be no impetus for k (or y) to change.

Notice second that if we observe an economy at a level of k other than k^s, economic forces will be driving k toward k^s. For example, consider an economy with a level of k below k^s. For such an economy, $sf(k)$ would be greater than $k(n+d)$, and $\dot{k}$ would be positive. This means that k is growing and the economy is moving to the right in the diagram. k will continue to grow until it reaches the level k^s, at which point it ceases to grow. Similarly, if an economy started out with a level of k to the right of k^s the capital-labor ratio would be falling, and it would continue to fall until the economy finds itself in steady-state or long-run equilibrium at k^s.

Figure 4.2b demonstrates what would happen if we took an economy in initial steady-state equilibrium at k^1 and increased its saving rate of s^1 to s^2. The $sf(k)$ function would pivot up, remaining connected to the origin, as the unchanging $f(k)$ function is multiplied by a larger fraction. With an initial capital-labor ratio of k^1 and a new higher saving rate s^2, $s^2 f(k^1)$ would exceed $k^1(n+d)$, and $\dot{k}$ (and the rate of growth in output per worker) would become positive, but this new growth in k and y would eventually die out as the economy tends toward a new long-run equilibrium with zero growth in per capita income. The increased frugality would not be without benefit, because the economy would eventually find itself in a new steady-state equilibrium characterized by a higher capital-labor ratio, k^2, and higher income per capita (at the intersection between the $k(n+d)$ line and the higher $s^2 f(k)$ curve), but the sustained increase in s would produce a period of growth in output per worker that lasts only as long as it takes to achieve the new long-run equilibrium ratio of capital to labor k^2.

[2] This is true as long as the $sf(k)$ curve is not so low that it lies entirely beneath the $k(n+d)$ locus and that it is sufficiently curved that its slope does eventually fall below the slope of the $k(n+d)$ locus.

4.3C The neoclassical growth model with technical change

The implication that growth driven only by a constant saving rate should grind to a halt flies in the face of real-world experiences, such as that of the United States, which has managed to save and experience per capita income growth at fairly steady rates over long periods of time. Is this the sort of unfortunate theoretical implication that should lead us to discard the model? Solow's answer was that we should modify rather than discard the model. In particular, he suggested replacing the aggregate production function (equation 4.1) by the function

$$Y = F(K, AL) \tag{4.12}$$

where A is a parameter describing the level of technology. More specifically, A is a parameter whose increase signifies **labor augmenting technical change**. We can think of AL as the quantity of "effective labor" in the economy (i.e., the economy's capacity to fulfill labor tasks), and we can think of technical change (denoted by an increase in A) as change that allows each physical unit of labor to supply more effective labor. In the presence of labor-augmenting technical change, the ratio of capital to effective labor, and thus the marginal product of capital, can remain constant, even while the ratio of capital to physical labor is rising. Such technical change serves to counteract the diminishing returns to capital that caused growth to grind to a halt in the simplest neoclassical model.

To understand growth and long-run equilibrium in this economy, it is useful to define and understand the evolution over time of the ratio of capital to effective labor, K/AL, which we will call k^*. It can be shown that the growth rate of k^* is equal to

$$\frac{\dot{k}^*}{k^*} = \frac{\dot{K}}{K} - \frac{\dot{L}}{L} - \frac{\dot{A}}{A} \tag{4.13}$$

Making the same assumptions as before regarding the growth of K and L, assuming that A increases at the rate g, and multiplying both sides by k^*, this becomes

$$\dot{k}^* = sf(k^*) - k^*(n + d + g) \tag{4.14}$$

To understand the growth performance of the economy described by this equation, we turn to Figure 4.3. Here the horizontal axis measures the level of k^*, and the vertical axis measures output per effective worker ($y^* = Y/AL$) and investment per effective worker. Using the same logic as we applied to the previous diagram, we find that this economy will converge toward a steady-state equilibrium in which k^* is constant at the level k^{*s}, associated with the intersection of the two curves. Notice, however, that a steady state in which $k^* = K/AL$ is constant is a steady state in which K/L and GDP per capita must be rising! In fact, with A rising at the rate g, K/L must be rising at the constant rate g as well to hold k^* constant, and output per capita must be growing at the rate g, too. This model thus predicts that the economy will achieve a steady state in which output per worker grows at a constant rate, equal to the rate of labor-augmenting technical change, g. The introduction of technical change allows the model to explain sustained growth in the long run. While the rate of saving s does not influence the rate of growth in steady state, it continues to play a role in determining the steady-state level of k^* and thus the steady-state path of GDP per capita.

Solow's main purpose in constructing the neoclassical model with technical change was to construct a model that can explain the coexistence in a single country of a stable, positive growth rate and a stable rate of saving over a long period. But many subsequent researchers have pointed out the model's implications for differences across countries in rates of economic growth and have subjected these implications to empirical testing.

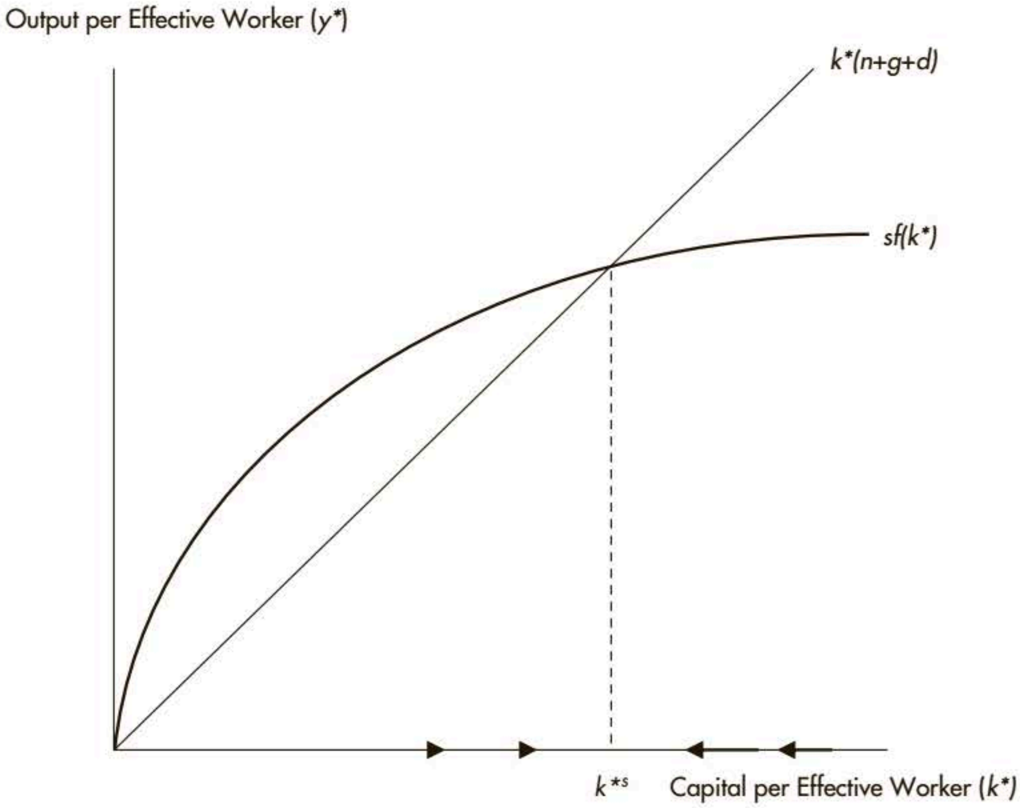

FIGURE 4.3
Steady-State Equilibrium in the Neoclassical Growth Model with Technical Change

According to the model, if all countries were already in steady state and had access to the same technology, then all countries should be growing at the same rate, g. Observing that growth rates differ greatly across countries, we must conclude that if this model is a good representation of reality, then countries must not be in steady state.

Among countries that are out of steady state but heading toward the same steady state (because they are characterized by the same values for s and $n+g+d$), the countries that start out poorer grow faster in this model. Diagrams like Figure 4.3 would be identical for such countries, except that poorer countries would be starting out farther to the left of the steady-state level of k^*. Thus the model implies that if all countries were heading toward the same steady state, we should see evidence of **unconditional convergence**, in which poorer countries (i.e., countries starting at lower levels of k^* and GDP per capita) simply grow faster than richer countries. Though these countries start out poorer, the model predicts that they will ultimately catch up, erasing those initial differences and enjoying the same level of growth and per capita income in steady state.

Unfortunately, the data graphed in Figure 3.1 contradict this rather optimistic prediction. In the world as a whole there has been no systematic tendency for poorer countries to grow faster than richer countries. Indeed, some of the poorest countries have also experienced the worst growth performance.

If countries are out of steady state and are heading toward different steady states (because they differ in their values of s and $n+g+d$), then among countries starting at the same level of k^* and income per capita, those with lower s and higher $n+g+d$ should be heading toward lower steady-state levels of k^*. Because they are closer to their ultimate steady states, they should be growing slower. Once we allow for these differences in the steady states toward which countries are heading, the model predicts **conditional convergence**. This means that even though poor countries don't always grow faster than richer countries, it may still be the case that after controlling for differences across countries in s and $n+g+d$, poorer countries grow faster than richer countries. That is, if you ran a regression of growth rates on measures of s, $n+g+d$ and initial income per capita, the coefficient on income per capita would be negative (and the coefficient on s would be positive and the coefficient on $n+g+d$ would be negative).

Regressions employing data from the last four decades often reveal such patterns of conditional convergence. Such patterns have sometimes been taken as evidence in favor of the neoclassical model and against alternative models proposed later (which we discuss below), but it is now broadly recognized that empirical conditional convergence results can be reconciled with a wide range of theoretical models and thus do not, in fact, help establish the superiority of the neoclassical model (Romer, 1994; Durlauf et al., 2008).

More elaborate versions of the neoclassical growth model, which pay more explicit attention to the choices of the economy's many producers and owners of labor and capital, reveal that as long as the technologies faced by individual producers are characterized by constant returns to scale, and as long as producers operate in perfectly competitive output and input markets, then the economy will deliver optimal growth and continuous full employment without the need for government intervention. People who care about the future and who operate in perfect financial markets will correctly perceive the incentive to save and invest. There is thus no need to use government policy to encourage more investment or production. In fact, by distorting incentives for factor use and investment, government intervention is more likely to hurt growth than to help it. The model is thus consistent with a world view of rising importance in the 1960s and 1970s, in which the key to speeding growth was to pull government out of production, reduce government intervention in markets, and reduce government efforts to protect specific sectors of the economy from international competition.

4.3D Neoclassical growth with human capital

Though the notion of human capital is often traced back to Adam Smith, it was not until the 1960s that research into the role of human capital in determining productivity and incomes really took off, encouraged by the rising availability of large individual-level datasets on schooling and earnings (at least in the United States), as well as the introduction of computing power adequate for econometric analysis of those new datasets (Mincer, 1970). The emerging literature quickly demonstrated an important role for schooling and experience in raising people's incomes, presumably by increasing their productivity as workers. It was natural, then, for researchers whose interest in economic growth was revived by Solow's theoretical contribution to consider augmenting the model to incorporate human capital as an additional determinant of aggregate production. One example of how human capital may be introduced formally into the Solow framework is found in Mankiw et al. (1992), who replace Solow's aggregate production function by

$$Y = F(K, H, AL) \tag{4.15}$$

where H is total human capital, A continues to represent labor-augmenting technical change, and $F(.)$ exhibits diminishing marginal returns in each of its three arguments.[3] The authors assume that the creation of physical and human capital is financed by saving out of income, and they assume that a fraction s of income is devoted to physical capital investment and a fraction b is devoted to human capital investment. As the reader might guess, in this augmented Solow model, the economy converges to a steady state in which the values of physical capital per effective worker $(k^* = K/AL)$ and human capital per effective worker $(h^* = H/AL)$ are constant, the steady-state levels of these variables depend on the investment rates (s and b) as well as $n + g + d$, and steady-state income per worker grows at the rate of technical progress.

While interest in human capital introduced only minor modifications to economic growth theory in the 1960s and 1970s, it was associated with dramatic changes in the practice of

[3] In fact, they assumed a particular functional form, $Y = K^{\alpha} H^{\beta} (AL)^{1-\alpha-\beta}$, which allowed them to derive equations defining steady-state per capita income as an explicit function of $(n + g + d)$, s, b, α, and β, as well as a term capturing the fact that in steady state, income per capita is continuously rising at the rate of technical change, g. They show that this equation does a surprisingly good job of explaining differences across countries in per capita income.

development work by the large international organizations, for whom investment in education became the centerpiece of many efforts to promote growth and development. The popularity of investment in education can be explained in part by strong suspicions that such investments would be simultaneously beneficial for promoting growth, reducing inequality, and meeting the basic needs of the poor.

4.4 The 1980s to the Present: Market Imperfections, Increasing Returns, and Poverty Traps

4.4A Themes

The late 1980s brought a resurgence of interest in the theoretical study of economic growth as new data brought to light new patterns of growth experiences across countries and as developments in theoretical economics provided new analytical tools (Romer, 1994). While study of growth during the 1960s and 1970s was shaped by a single theoretical framework (the neoclassical model), theorists in more recent decades have developed a plethora of new models as they seek to wrestle with two sets of questions that were raised but not answered by the neoclassical model:

- What drives technical change?
- If investment in physical capital, human capital, and technical change is so important for growth and future well-being, and if people care about the future, then why don't we see more investment in poor countries? That is, why are some countries stagnating with low rates of saving, investment, and growth?

The new models of this period have brought thinking about economic growth even farther from capital fundamentalism. While still acknowledging the importance of physical capital, they reinforce interest in human capital, expand interest in the creation and adoption of improved technologies, and introduce new interest in reducing inefficiency and waste.

The new models also offer more nuanced alternatives to the structuralism and dualism of the 1940s and 1950s than was offered by the extreme neoclassical vision of perfect markets and perfectly integrated sectors. The models highlight various reasons why reasonably well-functioning markets might fail to deliver ideal investment, technical change, and growth. Some models raise the possibility of a new sort of dualism, in which entire poor countries, or poor populations within countries, might find themselves trapped in using traditional, low-productivity production methods, even while their higher-income neighbors pursue modern, higher-productivity methods and enjoy greater growth and prosperity. Under such circumstances policy and institutions might have roles to play in encouraging investment, innovation, and growth.

Altogether, the growth models of the 1980s to the present encourage careful micro study of the decisions, markets, and institutions affecting investment and TFP growth, and they point to a variety of reasons why markets and institutions can fail to deliver ideal growth outcomes.

4.4B Endogenizing technical change

The primary objective of a first subset of new growth theory was to endogenize technical change. Theorists in this literature agreed with Solow that technical change is important for long-run growth, but they were unsatisfied with the way Solow simply assumed a constant rate of technical change that was determined outside the model. **Endogenizing the rate of technical change** requires identifying specific choices people make that might give rise to technical