Our main study begins with these building blocks already in place and focuses on the contributions in the neoclassical tradition since the late 1950s. We use the neoclassical methodology and language and rely on concepts such as aggregate capital stocks, aggregate production functions, and utility functions for representative consumers (who often have infinite horizons). We also use modern mathematical methods of dynamic optimization and differential equations. These tools, which are described in an appendix at the end of this book, are familiar today to most first-year graduate students in economics.

From a chronological viewpoint, the starting point for modern growth theory is the classic article of Ramsey (1928), a work that was several decades ahead of its time. Ramsey's treatment of household optimization over time goes far beyond its application to growth theory; it is hard now to discuss consumption theory, asset pricing, or even business-cycle theory without invoking the optimality conditions that Ramsey (and Fisher [1930]) introduced to economists. Ramsey's intertemporally separable utility function is as widely used today as the Cobb–Douglas production function. The economics profession did not, however, accept or widely use Ramsey's approach until the 1960s.

Between Ramsey and the late 1950s, Harrod (1939) and Domar (1946) attempted to integrate Keynesian analysis with elements of economic growth. They used production functions with little substitutability among the inputs to argue that the capitalist system is inherently unstable. Since they wrote during or immediately after the Great Depression, these arguments were received sympathetically by many economists. Although these contributions triggered a good deal of research at the time, very little of this analysis plays a role in today's thinking.

The next and more important contributions were those of Solow (1956) and Swan (1956). The key aspect of the Solow–Swan model is the neoclassical form of the production function, a specification that assumes constant returns to scale, diminishing returns to each input, and some positive and smooth elasticity of substitution between the inputs. This production function is combined with a constant-saving-rate rule to generate an extremely simple general-equilibrium model of the economy.

One prediction from these models, which has been exploited seriously as an empirical hypothesis only in recent years, is conditional convergence. The lower the starting level of real per capita GDP, relative to the long-run or steady-state position, the faster is the growth rate. This property derives from the assumption of diminishing returns to capital; economies that have less capital per worker (relative to their long-run capital per worker) tend to have higher rates of return and higher growth rates. The convergence is conditional because the steady-state levels of capital and output per worker depend, in the Solow–Swan model, on the saving rate, the growth rate of population, and the position of the production function—characteristics that might vary across economies. Recent empirical studies indicate that we should include additional sources of cross-country variation, especially differences in government policies and in initial stocks of human capital. The key point, however, is that the concept of conditional convergence—a basic property of the Solow-Swan model—has considerable explanatory power for economic growth across countries and regions.

Another prediction of the Solow–Swan model is that, in the absence of continuing improvements in technology, per capita growth must eventually cease. This prediction, which resembles those of Malthus and Ricardo, also comes from the assumption of diminishing returns to capital. We have already observed, however, that positive rates of per capita growth can persist over a century or more and that these growth rates have no clear tendency to decline.

The neoclassical growth theorists of the late 1950s and 1960s recognized this modeling deficiency and usually patched it up by assuming that technological progress occurred in an exogenous manner. This device can reconcile the theory with a positive, possibly constant per capita growth rate in the long run, while retaining the prediction of conditional convergence. The obvious shortcoming, however, is that the long-run per capita growth rate is determined entirely by an element—the rate of technological progress—that is outside of the model. (The long-run growth rate of the level of output also depends on the growth rate of population, another element that is exogenous in the standard theory.) Thus, we end up with a model of growth that explains everything but long-run growth, an obviously unsatisfactory situation.

Cass (1965) and Koopmans (1965) brought Ramsey's analysis of consumer optimization back into the neoclassical growth model and thereby provided for an endogenous determination of the saving rate. This extension allows for richer transitional dynamics but tends to preserve the hypothesis of conditional convergence. The endogeneity of saving also does not eliminate the dependence of the long-run per capita growth rate on exogenous technological progress.

The equilibrium of the Cass–Koopmans version of the neoclassical growth model can be supported by a decentralized, competitive framework in which the productive factors, labor and capital, are paid their marginal products. Total income then exhausts the total product because of the assumption that the production function features constant returns to scale. Moreover, the decentralized outcomes are Pareto optimal.

The inclusion of a theory of technological change in the neoclassical framework is difficult, because the standard competitive assumptions cannot be maintained. Technological advance involves the creation of new ideas, which are partially nonrival and therefore have aspects of public goods. For a given technology—that is, for a given state of knowledge—it is reasonable to assume constant returns to scale in the standard, rival factors of production, such as labor, capital, and land. In other words, given the level of knowledge on how to produce, one would think that it is possible to replicate a firm with the same amount of labor, capital, and land and obtain twice as much output. But then, the returns to scale tend to be increasing if the nonrival ideas are included as factors of production. These increasing returns conflict with perfect competition. In particular, the compensation of nonrival old ideas in accordance with their current marginal cost of production—zero—will not provide the appropriate reward for the research effort that underlies the creation of new ideas.

Arrow (1962) and Sheshinski (1967) constructed models in which ideas were unintended by-products of production or investment, a mechanism described as learning-by-doing. In these models, each person's discoveries immediately spill over to the entire economy, an instantaneous diffusion process that might be technically feasible because knowledge is nonrival. Romer (1986) showed later that the competitive framework can be retained in this case to determine an equilibrium rate of technological advance, but the resulting growth rate would typically not be Pareto optimal. More generally, the competitive framework breaks down if discoveries depend in part on purposive R&D effort and if an individual's innovations spread only gradually to other producers. In this realistic setting, a decentralized theory of technological progress requires basic changes in the neoclassical growth model to incorporate models of imperfect competition.[3] These additions to the theory did not come until Romer's (1987, 1990) research in the late 1980s.

The work of Cass (1965) and Koopmans (1965) completed the basic neoclassical growth model. Thereafter, growth theory became excessively technical and steadily lost contact with empirical applications. In contrast, development economists, who are required to give advice to sick countries, retained an applied perspective and tended to use models that were technically unsophisticated but empirically useful. The fields of economic development and economic growth drifted apart, and the two areas became almost completely separated.

Probably because of its lack of empirical relevance, growth theory effectively died as an active research field by the early 1970s, on the eve of the rational-expectations revolution and the oil shocks. For about 15 years, macroeconomic research focused on short-term fluctuations. Major contributions included the incorporation of rational expectations into business-cycle models, improved approaches to policy evaluation, and the application of general-equilibrium methods to real business-cycle theory.

Since the mid-1980s, research on economic growth has experienced a new boom, beginning with the work of Romer (1986) and Lucas (1988). The motivation for this research was the observation (or recollection) that the determinants of long-run economic growth are crucial issues, far more important than the mechanics of business cycles or the countercyclical effects of monetary and fiscal policies. But a recognition of the significance of long-run growth is only a first step; to go further, one has to escape the straitjacket of the neoclassical growth model, in which the long-term per capita growth rate is pegged by the rate of exogenous technological progress. Thus, in one way or another, the recent contributions determine the long-run growth rate within the model; hence, the designation *endogenous-growth* models.

The initial wave of the new research—Romer (1986), Lucas (1988), Rebelo (1991)—built on the work of Arrow (1962), Sheshinski (1967), and Uzawa (1965) and did not really introduce a theory of technological change. In these models, growth may go on indefinitely because the returns to investment in a broad class of capital goods—which includes human capital—do not necessarily diminish as economies develop. (This idea goes back to Knight [1944].) Spillovers of knowledge across producers and external benefits from human capital are parts of this process, but only because they help avoid the tendency for diminishing returns to the accumulation of capital.

The incorporation of R&D theories and imperfect competition into the growth framework began with Romer (1987, 1990) and includes significant contributions by Aghion and Howitt (1992) and Grossman and Helpman (1991, Chapters 3 and 4). In these models, technological advance results from purposive R&D activity, and this activity is rewarded by some form of *ex-post* monopoly power. If there is no tendency for the economy to run out of ideas, then the growth rate can remain positive in the long run. The rate of growth and the underlying amount of inventive activity tend, however, not to be Pareto optimal because of distortions related to the creation of the new goods and methods of production. In these frameworks, the long-term growth rate depends on governmental actions, such as taxation, maintenance of law and order, provision of infrastructure services, protection of intellectual property rights, and regulations of international trade, financial markets, and other aspects of the economy. The government therefore has great potential for good or ill through its influence on the long-term rate of growth.

[3]Another approach is to assume that all of the nonrival research—a classic public good—is financed by the government through involuntary taxes; see Shell (1967).

The new research also includes models of the diffusion of technology. Whereas the analysis of discovery relates to the rate of technological progress in leading-edge economies, the study of diffusion pertains to the manner in which follower economies share by imitation in these advances. Since imitation tends to be cheaper than innovation, the diffusion models predict a form of conditional convergence that resembles the predictions of the neoclassical growth model.

Another key exogenous parameter in the neoclassical growth model is the growth rate of population. A higher rate of population growth lowers the steady-state level of capital and output per worker and tends thereby to reduce the per capita growth rate for a given initial level of per capita output. The standard model does not, however, consider the effects of per capita income and wage rates on population growth—the kinds of effects stressed by Malthus—and also does not take account of the resources used up in the process of child rearing. Another line of recent research makes population growth endogenous by incorporating an analysis of fertility choice in the neoclassical model. The results are consistent, for example, with the empirical regularity that fertility rates tend to fall with per capita income over the main range of experience, but may rise with per capita income for the poorest countries. Additional work related to the endogeneity of labor supply in a growth context concerns migration and labor/leisure choice.

The clearest distinction between the growth theory of the 1960s and that of the 1980s and 1990s is that the recent research pays close attention to empirical implications and to the relation between theory and data. Some of this applied perspective involves amplification of the empirical implications of the older theory, notably the neoclassical growth model's prediction of conditional convergence. Other analyses apply more directly to the recent theories of endogenous growth, including the roles of increasing returns, R&D activity, human capital, and the diffusion of technology.

In this book we attempt to reflect the recent emphasis on the interplay between theory and applications. Thus, we stress the empirical implications of the various theories that we develop. We also include three chapters that are devoted entirely to data and empirical analyses.

The recent growth research has attracted interest from economists in a wide variety of fields. Conferences on growth have participation from specialists in macroeconomics, development, international economics, theory, history, econometrics, and industrial organization. We think that the effective combination of theory and empirical work will sustain this broad appeal and will allow growth theory to survive this time as a vibrant field. We do not expect the growth theory of the 1990s to suffer the same fate as the growth theory of the 1960s.

CHAPTER 1

GROWTH MODELS WITH EXOGENOUS SAVING RATES (THE SOLOW–SWAN MODEL)

1.1 THE BASIC STRUCTURE

All the models of growth that we discuss in this book have the same basic general-equilibrium structure. First, households (or families) own the inputs and assets of the economy, including ownership rights in firms, and choose the fractions of their income to consume and save. Each household determines how many children to have, whether to join the labor force, and how much to work. Second, firms hire inputs, such as capital and labor, and use these inputs to produce goods that they sell to households or other firms. Firms have access to a technology—which may evolve over time—that allows them to transform inputs into output. Third, markets exist on which firms sell goods to households or other firms and on which households sell the inputs to firms. The quantities demanded and supplied determine the relative prices of the inputs and the produced goods.

It is convenient in this initial chapter to use a simplified setup that excludes markets and firms. We can think of a composite unit—a household/producer like Robinson Crusoe—who owns the inputs and also manages the technology that transforms inputs into outputs. There are only two inputs, physical capital, $K(t)$, and labor, $L(t)$. The production function takes the form

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

where $Y(t)$ is the flow of output produced at time t. The production function depends on time, t, to reflect the effects of technological progress: the same amount of capital and labor yields a larger quantity of output in 1995 than in 1895 if the technology employed in 1995 is superior.

We assume a one-sector production technology in which output is a homogeneous good that can be consumed, $C(t)$, or invested, $I(t)$, to create new units of physical capital, $K(t)$. One way to think about the one-sector technology is to draw an analogy with farm animals, which can be eaten or used as inputs to produce more farm animals. The literature on economic growth has used more inventive examples—with such terms as *shmoos*, *putty*, or *ectoplasm*—to reflect the easy transmutation of capital goods into consumables, and vice versa.

We assume in this chapter that the economy is closed: households cannot buy foreign goods or assets and cannot sell home goods or assets abroad. (Chapter 3 allows for an open economy.) In a closed economy, output equals income, and the amount invested equals the amount saved.

Let $s(\cdot)$ be the fraction of output that is saved—that is, the *saving rate*—so that $1 - s(\cdot)$ is the fraction of output that is consumed. Rational households choose the saving rate by comparing the costs and benefits of consuming today rather than tomorrow; this comparison involves preference parameters and variables that describe the state of the economy, such as the level of wealth and the interest rate. In Chapter 2, where we model this decision explicitly, we find that $s(\cdot)$ is a complicated function for which there are typically no closed-form solutions. To facilitate the analysis in this initial chapter, we assume that $s(\cdot)$ is given exogenously. The simplest function, the one assumed by Solow (1956) and Swan (1956) in their classic articles, is a constant, $s(\cdot) = s > 0$. We use this constant-saving-rate specification in this chapter because it brings out a large number of results in a clear way.

We assume that capital depreciates at the constant rate $\delta > 0$; that is, at each point in time, a constant fraction of the capital stock wears out and, hence, can no longer be used for production. (If we think of goods as farm animals, then a constant fraction of the animals dies at each moment, unrealistically independent of the average age of the stock.)

The net increase in the stock of physical capital at a point in time equals gross investment less depreciation:

$$\dot{K} = I - \delta K = s \cdot F(K, L, t) - \delta K, \tag{1.2}$$

where a dot over a variable, such as $\dot{K}$, denotes differentiation with respect to time, and $0 \leq s \leq 1$. Equation (1.2) determines the dynamics of K for a given technology and labor force. In the first sections of this chapter, we neglect technological progress; that is, we assume that $F(\cdot)$ is independent of t. This assumption will be relaxed later.

The labor force, L, varies over time because of population growth, changes in participation rates, and shifts in the amount of time worked by the typical worker. The growth of population reflects, in turn, the behavior of fertility, mortality, and migration. Chapter 9 allows for choices between work and leisure and also considers the effects from migration, fertility, and mortality on population. In this chapter, we simplify by assuming that population grows at a constant, exogenous rate, $\dot{L}/L = n \geq 0$, and that everyone works at a given intensity. If we normalize the number of people at time 0 to 1 and the work intensity per person also to 1, then the population and labor force at time t are equal to

$$L(t) = e^{nt}. \tag{1.3}$$

If $L(t)$ is given from Eq. (1.3) and technological progress is absent, then Eq. (1.2) determines the time paths of capital, K, and output, Y. In the next sections, we show that this behavior depends crucially on the properties of the production function, $F(\bullet)$. In fact, apparently minor differences in assumptions about $F(\bullet)$ can generate radically different theories of economic growth.

1.2 THE NEOCLASSICAL MODEL OF SOLOW AND SWAN

1.2.1 The Neoclassical Production Function

If we neglect technological progress, then the production function from Eq. (1.1) takes the form

$$Y = F(K, L). \tag{1.4}$$

We say that the production function is *neoclassical* if the following three properties are satisfied. First, for all $K > 0$ and $L > 0$, $F(\bullet)$ exhibits positive and diminishing marginal products with respect to each input:

$$\frac{\partial F}{\partial K} > 0, \quad \frac{\partial^2 F}{\partial K^2} < 0 \qquad \frac{\partial F}{\partial L} > 0, \quad \frac{\partial^2 F}{\partial L^2} < 0. \tag{1.5a}$$

Second, $F(\bullet)$ exhibits constant returns to scale:

$$F(\lambda K, \lambda L) = \lambda \cdot F(K, L) \text{ for all } \lambda > 0. \tag{1.5b}$$

Third, the marginal product of capital (or labor) approaches infinity as capital (or labor) goes to 0 and approaches 0 as capital (or labor) goes to infinity:

$$\lim_{K\to 0}(F_K) = \lim_{L\to 0}(F_L) = \infty \qquad \lim_{K\to \infty}(F_K) = \lim_{L\to \infty}(F_L) = 0 \tag{1.5c}$$

These last properties are called *Inada conditions*, following Inada (1963).

The condition of constant returns to scale implies that output can be written as

$$Y = F(K, L) = L \cdot F(K/L, 1) = L \cdot f(k),$$

where $k \equiv K/L$ is the capital–labor ratio, $y \equiv Y/L$ is per capita output, and the function $f(k)$ is defined to equal $F(k, 1)$. This result means that the production function can be expressed in *intensive form* as

$$y = f(k). \tag{1.6}$$

We can use the condition $Y = L \cdot f(k)$ and differentiate with respect to K, for fixed L, and then with respect to L, for fixed K, to verify that the marginal products of the factor inputs are given by

$$\partial Y/\partial K = f'(k), \qquad \partial Y/\partial L = [f(k) - k \cdot f'(k)]. \tag{1.7}$$

The Inada conditions imply $\lim_{k\to 0}[f'(k)] = \infty$ and $\lim_{k\to\infty}[f'(k)] = 0$.

We can show that the neoclassical properties, Eqs. (1.5a)–(1.5c), imply that each input is essential for production, that is, $F(0, L) = F(K, 0) = f(0) = 0$. The properties also imply that output goes to infinity as either input goes to infinity. See the appendix at the end of this chapter for proofs of these propositions.

One simple production function that is often thought to provide a reasonable description of actual economies is the Cobb–Douglas function,

$$Y = AK^{\alpha}L^{1-\alpha}, \tag{1.8}$$

where $A > 0$ is the level of the technology, and α is a constant with $0 < \alpha < 1$. The Cobb–Douglas function can be written in intensive form as

$$y = Ak^{\alpha}. \tag{1.9}$$

Note that $f'(k) = A\alpha k^{\alpha-1} > 0$, $f''(k) = -A\alpha(1-\alpha)k^{\alpha-2} < 0$, $\lim_{k\to\infty} f'(k) = 0$, and $\lim_{k\to 0} f'(k) = \infty$. Thus, the Cobb–Douglas form satisfies the properties of a neoclassical production function.

1.2.2 The Fundamental Dynamic Equation for the Capital Stock

We now analyze the dynamic behavior of the economy described by the neoclassical production function. The resulting growth model is called the Solow–Swan model, after the important contributions of Solow (1956) and Swan (1956).

The change in the capital stock over time is given by Eq. (1.2). If we divide both sides of this equation by L, then we get

$$\dot{K}/L = s \cdot f(k) - \delta k.$$

The right-hand side contains per capita variables only, but the left-hand side does not. We can write $\dot{K}/L$, as a function of k by using the condition

$$\dot{k} \equiv \frac{d(K/L)}{dt} = \dot{K}/L - nk,$$