Economic Growth

Second Edition

Robert J. Barro
Xavier Sala-i-Martin

The MIT Press
Cambridge, Massachusetts
London, England

This book was set in Times Roman by ICC Typesetting and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Barro, Robert J.
Economic growth / Robert J. Barro, Xavier Sala-i-Martin—2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-262-02553-1
1. Economic development—Mathematical models. I. Sala-i-Martin, Xavier. II. Title.

HD75.5.B37 2003
338.9′001′51—dc22

2003059392

To Rachel
—*Robert J. Barro*

A la memòria dels meus estimats Joan Martín Pujol i Ramon Oriol Martín Montemayor
—*Xavier Sala-i-Martin*

Contents

Preface

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 about anything else.*[1]
—Robert E. Lucas, Jr. (1988)

Economists have, in some sense, always known that growth is important. Yet, at the core of the discipline, the study of economic growth languished after the late 1960s. Then, after a lapse of two decades, this research became vigorous again in the late 1980s. The new research began with models of the determination of long-run growth, an area that is now called endogenous growth theory. Other recent research extended the older, neoclassical growth model, especially to bring out the empirical implications for convergence across economies. This book combines new results with expositions of the main research that appeared from the 1950s through the beginning of the 2000s. The discussion stresses the empirical implications of the theories and the relation of these hypotheses to data and evidence. This combination of theory and empirical work is the most exciting aspect of ongoing research on economic growth.

The introduction motivates the study, brings out some key empirical regularities in the growth process, and provides a brief history of modern growth theory. Chapters 1 and 2 deal with the neoclassical growth model, from Solow–Swan in the 1950s, to Cass–Koopmans (and recollections of Ramsey) in the 1960s, to recent refinements of the model. Chapter 3 deals with extensions to incorporate a government sector and to allow for adjustment costs in investment, as well as with the open economy and finite-horizon models of households. Chapters 4 and 5 cover the versions of endogenous growth theory that rely on forms of constant returns to reproducible factors. Chapters 6, 7, and 8 explore recent models of technological change and R&D, including expansions in the variety and quality of products and the diffusion of knowledge. Chapter 9 allows for an endogenous determination of labor supply and population, including models of migration, fertility, and labor/leisure choice. Chapter 10 works out the essentials of growth accounting and applies this framework to the endogenous growth models. Chapter 11 covers empirical analysis of regions of countries, including the U.S. states and regions of Europe and Japan. Chapter 12 deals with empirical evidence on economic growth for a broad panel of countries from 1960 to 2000.

1. These inspirational words from Lucas have probably become the most frequently quoted passage in the growth literature. Thus it is ironic (and rarely mentioned) that, even while Lucas was writing his ideas, India had already begun to grow faster than Indonesia and Egypt. The growth rates of GDP per person from 1960 to 1980 were 3.2% per year in Egypt, 3.9% in Indonesia, and 1.5% in India. In contrast, from 1980 to 2000, the growth rates of GDP per person were 1.8% per year in Egypt, 3.5% in Indonesia, and 3.6% in India. Thus, the Indian government seems to have met Lucas's challenge, whereas Egypt was faltering.

The material is written as a text at the level of first-year graduate students in economics. The widely used first edition has proven successful for graduate courses in macroeconomics, economic growth, and economic development. Most of the chapters include problems that guide the students from routine exercises through suggestive extensions of the models. The level of mathematics includes differential equations and dynamic optimization, topics that are discussed in the mathematical appendix at the end of the book. For undergraduates who are comfortable with this level of mathematics, the book works well for advanced, elective courses. The first edition has been used at this level throughout the world.

We have benefited from comments by Daron Acemoglu, Philippe Aghion, Minna S. Andersen, Marios Angeletos, Elsa V. Artadi, Abhijit Banerjee, Paulo Barelli, Gary Becker, Olivier Blanchard, Juan Braun, Francesco Caselli, Paul Cashin, Daniel Cohen, Irwin Collier, Diego Comin, Michael Connolly, Michelle Connolly, Ana Corbacho, Vivek Dehejia, Marcelo Delajara, Gernot Doppelhoffer, Paul Evans, Rosa Fernandez, Monica Fuentes-Neira, Xavier Gabaix, Oded Galor, Victor Gomes Silva, Zvi Griliches, Gene Grossman, Christian Groth, Laila Haider, Elhanan Helpman, Toshi Ichida, Dale Jorgenson, Ken Judd, Jinill Kim, Michael Kremer, Phil Lane, Stephen Lin, Norman Loayza, Greg Mankiw, Kiminori Matsuyama, Sanket Mohapatra, Casey Mulligan, Kevin M. Murphy, Marco Neuhaus, Renger van Nieuwkoop, Sylvia Noin-McDavid, Joan O'Connell, Salvador Ortigueira, Lluis Parera, Pietro Peretto, Torsten Persson, Danny Quah, Climent Quintana, Rodney Ramchandran, Jordan Rappaport, Sergio Rebelo, Joan Ribas, Paul Romer, Joan Rossello, Michael Sarel, Etsuro Shioji, Chris Sims, B. Anna Sjögren, Nancy Stokey, Gustavo Suarez, Robert Tamura, Silvana Tenreyro, Merritt Tilney, Aaron Tornell, Nuri Ucar, Jaume Ventura, Martin Weitzman, Arthur Woll, and Alwyn Young.

About the Authors

Robert J. Barro is Robert C. Waggoner Professor of Economics at Harvard University. He has a B.S. in physics from Caltech and a Ph.D. in economics from Harvard, and he previously held faculty positions at Rochester, Chicago, and Brown. He is a viewpoint columnist for *Business Week,* a senior fellow of the Hoover Institution at Stanford, and a Research Associate of the National Bureau of Economic Research. In 2003 he was president-elect of the Western Economic Association, in 1997–98 he was vice president of the American Economic Association, and in 1994–95 he was Houblon-Norman Research Fellow at the Bank of England. He is married to Rachel McCleary, with whom he codirects the Project on Religion, Economy, and Society at Harvard University.

Xavier Sala-i-Martin is a professor of economics at Columbia University and a visiting professor at the Universitat Pompeu Fabra (in Barcelona). He has a B.S. from Universitat Autònoma de Barcelona and a Ph.D. from Harvard University. He is a Research Associate of the National Bureau of Economic Research and the Center for European Policy Research. He is a columnist for *La Vanguardia* in Barcelona and a contributor to a number of shows on Channel 3 of the Catalan Television and Catalunya Radio. He is a senior economic adviser to the World Economic Forum and a member of the board of the Fundacio Catalunya Oberta. In 1992, 1995, 1998, and 1999, the students at Yale and Columbia honored him with the Distinguished Teacher Award for his classes on economic growth.

Introduction

I.1 The Importance of Growth

To think about the importance of economic growth, we begin by assessing the long-term performance of the U.S. economy. The real per capita gross domestic product (GDP) in the United States grew by a factor of 10 from $3340 in 1870 to $33,330 in 2000, all measured in 1996 dollars. This increase in per capita GDP corresponds to a growth rate of 1.8 percent per year. This performance gave the United States the second-highest level of per capita GDP in the world in 2000 (after Luxembourg, a country with a population of only about 400,000).[1]

To appreciate the consequences of apparently small differentials in growth rates when compounded over long periods of time, we can calculate where the United States would have been in 2000 if it had grown since 1870 at 0.8 percent per year, one percentage point per year below its actual rate. A growth rate of 0.8 percent per year is close to the rate experienced in the long run—from 1900 to 1987—by India (0.64 percent per year), Pakistan (0.88 percent per year), and the Philippines (0.86 percent per year). If the United States had begun in 1870 at a real per capita GDP of $3340 and had then grown at 0.8 percent per year over the next 130 years, its per capita GDP in 2000 would have been $9450, only 2.8 times the value in 1870 and 28 percent of the actual value in 2000 of $33,330. Then, instead of ranking second in the world in 2000, the United States would have ranked 45th out of 150 countries with data. To put it another way, if the growth rate had been lower by just 1 percentage point per year, the U.S. per capita GDP in 2000 would have been close to that in Mexico and Poland.

Suppose, alternatively, that the U.S. real per capita GDP had grown since 1870 at 2.8 percent per year, 1 percentage point per year greater than the actual value. This higher growth rate is close to those experienced in the long run by Japan (2.95 percent per year from 1890 to 1990) and Taiwan (2.75 percent per year from 1900 to 1987). If the United States had still begun in 1870 at a per capita GDP of $3340 and had then grown at 2.8 percent per year over the next 130 years, its per capita GDP in 2000 would have been $127,000—38 times the value in 1870 and 3.8 times the actual value in 2000 of $33,330. A per capita GDP of $127,000 is well outside the historical experience of any country and may, in fact, be infeasible (although people in 1870 probably would have thought the same about $33,330). We can say, however, that a continuation of the long-term U.S. growth rate of 1.8 percent per year implies that the United States will not attain a per capita GDP of $127,000 until 2074.

1. The long-term data on GDP come from Maddison (1991) and are discussed in chapter 12. Recent data are from Heston, Summers, and Aten (2002) and are also discussed in chapter 12.

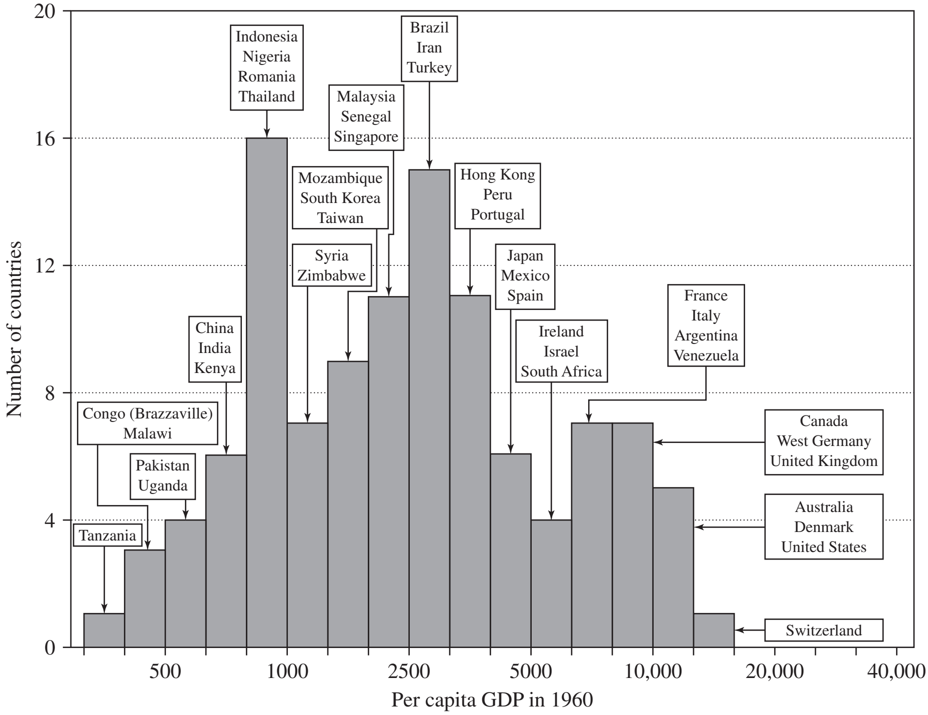

Figure I.1
Histogram for per capita GDP in 1960. The data, for 113 countries, are the purchasing-power-parity (PPP) adjusted values from Penn World Tables version 6.1, as described in Summers and Heston (1991) and Heston, Summers, and Aten (2002). Representative countries are labeled within each group.

The comparison of levels of real per capita GDP over a century involves multiples as high as 20; for example, Japan's per capita GDP in 1990 was about 20 times that in 1890. Comparisons of levels of per capita GDP across countries at a point in time exhibit even greater multiples. Figure I.1 shows a histogram for the log of real per capita GDP for 113 countries (those with the available data) in 1960. The mean value corresponds to a per capita GDP of $3390 (1996 U.S. dollars). The standard deviation of the log of real per capita GDP—a measure of the proportionate dispersion of real per capita GDP—was 0.89. This number means that a 1-standard-deviation band around the mean encompassed a range from 0.41 of the mean to 2.4 times the mean. The highest per capita GDP of $14,980 for Switzerland was 39 times the lowest value of $381 for Tanzania. The United States was second with a value of $12,270. The figure shows representative countries for each range of per capita GDP. The broad picture is that the richest countries included the OECD and

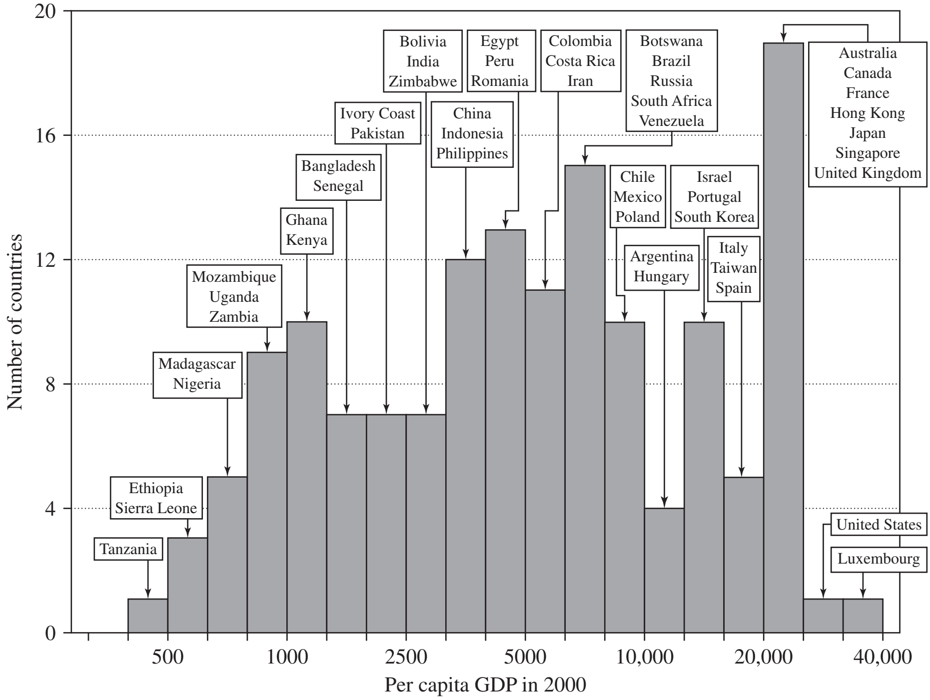

Figure I.2
Histogram for per capita GDP in 2000. The data, for 150 countries, are from the sources noted for figure I.1. Representative countries are labeled within each group.

a few places in Latin America, such as Argentina and Venezuela. Most of Latin America was in a middle range of per capita GDP. The poorer countries were a mixture of African and Asian countries, but some Asian countries were in a middle range of per capita GDP.

Figure I.2 shows a comparable histogram for 150 countries in 2000. The mean here corresponds to a per capita GDP of $8490, 2.5 times the value in 1960. The standard deviation of the log of per capita GDP in 2000 was 1.12, implying that a 1-standard-deviation band ranged from 0.33 of the mean to 3.1 times the mean. Hence, the proportionate dispersion of per capita GDP increased from 1960 to 2000. The highest value in 2000, $43,990 for Luxembourg, was 91 times the lowest value—$482 for Tanzania. (The Democratic Republic of Congo would be poorer, but the data are unavailable for 2000.) If we ignore Luxembourg because of its small size and compare Tanzania's per capita GDP with the second-highest value, $33,330 for the United States, the multiple is 69. Figure I.2 again marks out representative countries within each range of per capita GDP. The OECD countries still dominated the top group, joined by some East Asian countries. Most other Asian countries were in the middle range of per capita GDP, as were most Latin American countries. The lower range in 2000 was dominated by sub-Saharan Africa.

To appreciate the spreads in per capita GDP that prevailed in 2000, consider the situation of Tanzania, the poorest country shown in figure I.2. If Tanzania were to grow at the long-term U.S. rate of 1.8 percent per year, it would take 235 years to reach the 2000 level of U.S. per capita GDP. The required interval would still be 154 years if Tanzania were to grow at the long-term Japanese rate of 2.75 percent per year.

For 112 countries with the necessary data, the average growth rate of real per capita GDP between 1960 and 2000 was 1.8 percent per year—coincidentally the same as the long-term U.S. rate—with a standard deviation of 1.7.[2] Figure I.3 has a histogram of these growth rates; the range is from −3.2 percent per year for the Democratic Republic of Congo (the former Zaire) to 6.4 percent per year for Taiwan. (If not for missing data, the lowest-growing country would probably be Iraq.) Forty-year differences in growth rates of this magnitude have enormous consequences for standards of living. Taiwan raised its real per capita GDP by a factor of 13 from $1430 in 1960 (rank 76 out of 113 countries) to $18,730 in 2000 (rank 24 of 150), while the Democratic Republic of Congo lowered its real per capita GDP by a factor of 0.3 from $980 in 1960 (rank 93 of 113) to $320 in 1995—if not for missing data, this country would have the lowest per capita GDP in 2000.

A few other countries had growth rates from 1960 to 2000 that were nearly as high as Taiwan's; those with rates above 5 percent per year were Singapore with 6.2 percent, South Korea with 5.9 percent, Hong Kong with 5.4 percent, and Botswana with 5.1 percent. These countries increased their levels of per capita GDP by a multiple of at least 7 over 40 years. Just below came Thailand and Cyprus at 4.6 percent growth, China at 4.3 percent, Japan at 4.2 percent (with rapid growth mainly into the 1970s), and Ireland at 4.1 percent. Figure I.3 shows that a number of other OECD countries came in the next-highest growth groups, along with a few countries in Latin America (including Brazil and Chile) and more in Asia (including Indonesia, India, Pakistan, and Turkey). The United States ranked 40th in growth with a rate of 2.5 percent.

At the low end of growth, 16 countries aside from the Democratic Republic of Congo had negative growth rates of real per capita GDP from 1960 to 2000. The list (which would be substantially larger if not for missing data), starting from the bottom, is Central African Republic, Niger, Angola, Nicaragua, Mozambique, Madagascar, Nigeria, Zambia,

2. These statistics include the Democratic Republic of Congo (the former Zaire), for which the data are for 1960 to 1995.

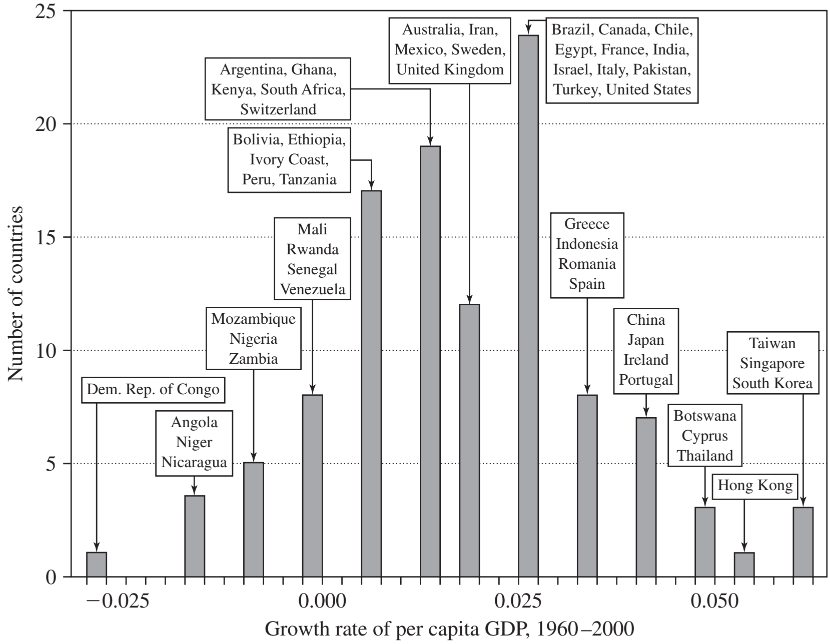

Figure I.3
Histogram for growth rate of per capita GDP from 1960 to 2000. The growth rates are computed for 112 countries from the values of per capita GDP shown for 1960 and 2000 in figures I.1 and I.2. For Democratic Republic of Congo (former Zaire), the growth rate is for 1960 to 1995. West Germany is the only country included in figure I.1 (for 1960) but excluded from figure I.3 (because of data problems caused by the reunification of Germany). Representative countries are labeled within each group.

Chad, Comoros, Venezuela, Senegal, Rwanda, Togo, Burundi, and Mali. Thus, except for Nicaragua and Venezuela, this group comprises only sub-Saharan African countries. For the 38 sub-Saharan African countries with data, the mean growth rate from 1960 to 2000 was only 0.6 percent per year. Hence, the typical country in sub-Saharan Africa increased its per capita GDP by a factor of only 1.3 over 40 years. Just above the African growth rates came a few slow-growing countries in Latin America, including Bolivia, Peru, and Argentina.

As a rough generalization for regional growth experiences, we can say that sub-Saharan Africa started relatively poor in 1960 and grew at the lowest rate, so it ended up by far the poorest area in 2000. Asia started only slightly above Africa in many cases but grew rapidly and ended up mostly in the middle. Latin America started in the mid to high range, grew somewhat below average, and therefore ended up mostly in the middle along with Asia.

Finally, the OECD countries started highest in 1960, grew in a middle range or better, and therefore ended up still the richest.

If we want to understand why countries differ dramatically in standards of living (figures I.1 and I.2), we have to understand why countries experience such sharp divergences in long-term growth rates (figure I.3). Even small differences in these growth rates, when cumulated over 40 years or more, have much greater consequences for standards of living than the kinds of short-term business fluctuations that have typically occupied most of the attention of macroeconomists. To put it another way, if we can learn about government policy options that have even small effects on long-term growth rates, we can contribute much more to improvements in standards of living than has been provided by the entire history of macroeconomic analysis of countercyclical policy and fine-tuning. Economic growth—the subject matter of this book—is the part of macroeconomics that really matters.

I.2 The World Income Distribution

Although we focus in this book on the theoretical and empirical determinants of aggregate economic growth, we should keep in mind that growth has important implications for the welfare of individuals. In fact, aggregate growth is probably the single most important factor affecting individual levels of income. Hence, understanding the determinants of aggregate economic growth is the key to understanding how to increase the standards of living of individuals in the world and, thereby, to lessen world poverty.

Figure I.4 shows the evolution of the world's per capita GDP from 1970 to 2000.[3] It is clear that the average person on the planet has been getting richer over time. But the positive average growth rate over the last three decades does not mean that the income of all citizens has increased. In particular, it does not mean that the incomes of the poorest people have grown nor that the number of people whose incomes are below a certain poverty line (say one dollar a day, as defined by the World Bank) has declined.[4] Indeed, if inequality

3. The "world" is approximated by the 126 countries (139 countries after the breakup of the Soviet Union in 1989) in Sala-i-Martin (2003a, 2003b). The individuals in these 126 countries made up about 95 percent of the world's population. World GDP per capita is estimated by adding up the data for individual countries from Heston, Summers, and Aten (2002) and then dividing by the world's population.

4. The quest for a "true" poverty line has a long tradition, but the current "one-dollar-a-day" line can be traced back to World Bank (1990). The World Bank originally defined the poverty line as one dollar a day in 1985 prices. Although the World Bank's own definition later changed to 1.08 dollars a day in 1993 dollars (notice that one 1985 dollar does not correspond to 1.08 1993 dollars), we use the original definition of one dollar a day in 1985 prices. One dollar a day (or 365 dollars a year) in 1985 prices becomes 495 dollars per year in 1996 prices, which is the base year of the Heston, Summers, and Aten (2002) data used to construct the world income distributions. Following Bhalla (2002), Sala-i-Martin (2003a) adjusts this poverty line upward by 15 percent to correct for the bias generated by the underreporting of the rich. This adjustment means that our "one-dollar-a-day" poverty line represents 570 dollars a year (or 1.5 dollars a day) in 1996 dollars.

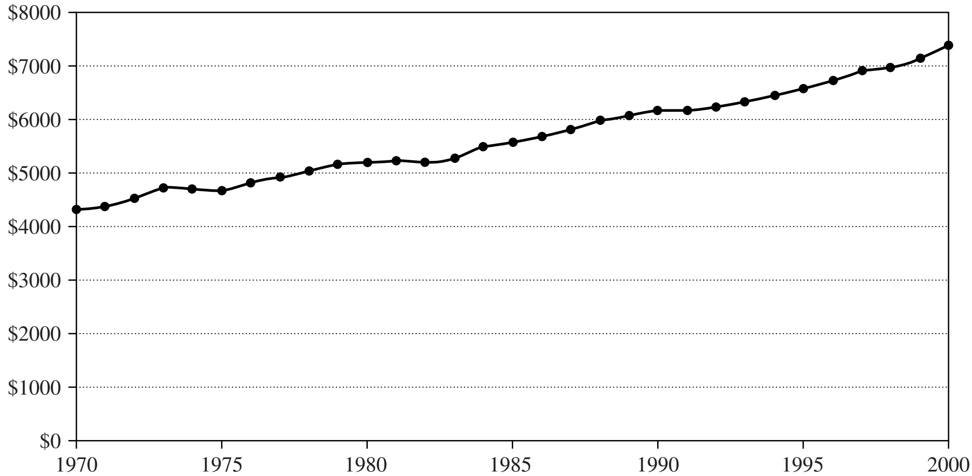

Figure I.4
World per capita GDP, 1970–2000. World per capita GDP is the sum of the GDPs for 126 countries (139 countries after the collapse of the Soviet Union) divided by population. The sample of 126 countries is the one used in Sala-i-Martin (2003a) and accounts for 95 percent of the world's population.

increased along with economic growth, it is possible for the world to have witnessed both positive per capita GDP growth and an increasing number of people below the poverty line. To assess how aggregate growth affects poverty, Sala-i-Martin (2003a) estimates the world distribution of individual income. To do so, he combines microeconomic survey and aggregate GDP data for each country, for every year between 1970 and 2000.[5] The result for 1970 is displayed in figure I.5. The horizontal axis plots the level of income (on a logarithmic scale), and the vertical axis has the number of people. The thin lines correspond to the income distributions of individual countries. Notice, for example, that China (the most populated country in the world) has a substantial fraction of the distribution below the $1/day line. The same is true for India and a large number of smaller countries. This pattern contrasts with the position of countries such as the United States, Japan, or even the USSR, which have very little of their distributions below the $1/day line. The thick line in figure I.5 is the integral of all the individual distributions. Therefore,

5. Sala-i-Martin (2003b) constructs an analogous distribution from which he estimates the number of people whose personal consumption expenditure is less than one dollar a day. The use of consumption, rather than income, accords better with the concept of "extreme poverty" used by international institutions such as the World Bank and the United Nations. However, personal consumption has the drawbacks of giving no credit to public services and saving.

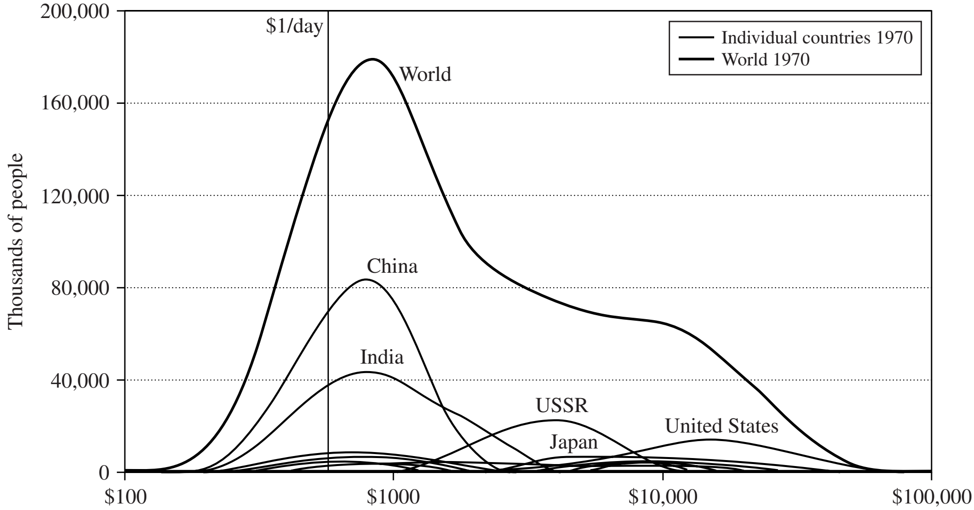

Figure I.5
The world distribution of income in 1970. The level of income is on the horizontal axis (on a logarithmic scale), and the number of people is on the vertical axis. The thin curves correspond to the income distributions of individual countries. The thick curve is the integral of individual country distributions and corresponds to the world distribution of income. The vertical line marks the poverty line (which corresponds to one dollar a day in 1985 prices). Source: Sala-i-Martin (2003a).

this line corresponds to the world distribution of income in 1970. Again, a substantial fraction of the world's citizens were poor (that is, had an income of less than $1/day) in 1970.

Figure I.6 displays the corresponding distributions for 2000. If one compares the 1970 with the 2000 distribution, one sees a number of interesting things. First, the world distribution of income has shifted to the right. This shift corresponds to the cumulated growth of per capita GDP. Second, we see that, underlying the evolution of worldwide income, there is a positive evolution of incomes in most countries in the world. Most countries increased their per capita GDP and, therefore, shifted to the right. Third, we see that the dispersion of the distributions for some countries, notably China, has increased over this period. In other words, income inequality rose within some large countries. Fourth, the increases in inequality within some countries have not been nearly enough to offset aggregate per capita growth, so that the fraction of the world's people whose incomes lie below the poverty line has declined dramatically.

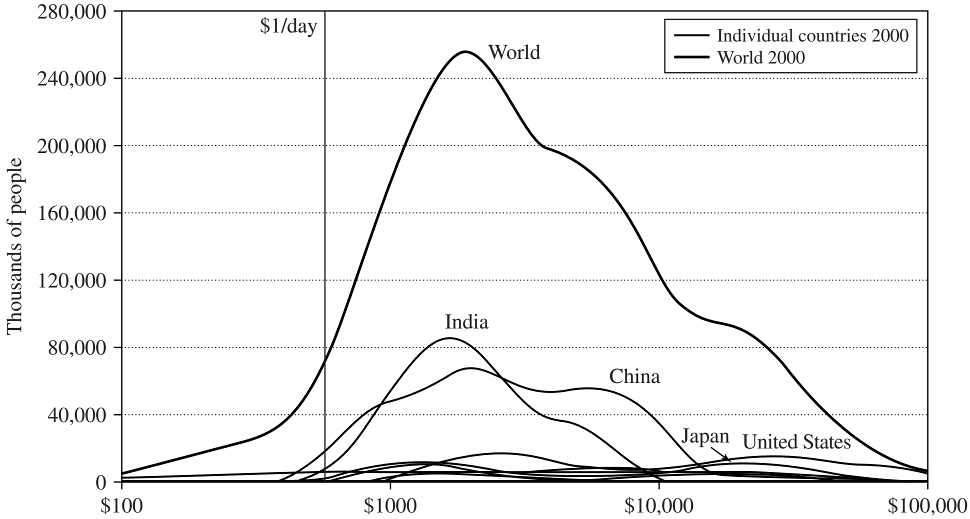

Figure I.6
The world distribution of income in 2000. The level of income is on the horizontal axis (on a logarithmic scale), and the number of people is on the vertical axis. The thin curves correspond to the income distributions of individual countries. The thick curve is the integral of individual country distributions and corresponds to the world distribution of income. The vertical line marks the poverty line (which corresponds to one dollar a day in 1985 prices). Source: Sala-i-Martin (2003a).

The exact fraction of the world's citizens that live below the poverty line can be computed from the distributions estimated by Sala-i-Martin (2003a).[6] These poverty rates, reported in figure I.7, have been cut by a factor of 3: whereas 20 percent of the world's citizens were poor in 1970, only 7 percent were poor in 2000.[7] Between 1970 and 1978, population growth more than offset the reduction in poverty rates. Indeed, Sala-i-Martin (2003a) shows that, during that period, the overall number of poor increased by 20 million people. But, since 1978, the total number of people with income below the $1/day threshold declined by more than 300 million. This achievement is all the more remarkable if we take into acount that overall population increased by more than 1.6 billion people during this period.

6. The World Bank, the United Nations, and many individual researchers define poverty in terms of consumption, rather than income. Sala-i-Martin (2003b) estimates poverty rates and head counts using consumption. The evolution of consumption poverty is similar to the one reported here for income although, obviously, the poverty rates are higher if one uses consumption instead of income and still uses the same poverty line.

7. Sala-i-Martin (2003a) reports cumulative distribution functions (CDFs) for 1970, 1980, 1990, and 2000. Using these CDFs, one can easily see that poverty rates have fallen dramatically over the last thirty years regardless of what poverty line one adopts. Thus, the conclusion that aggregate growth has reduced poverty is quite robust.

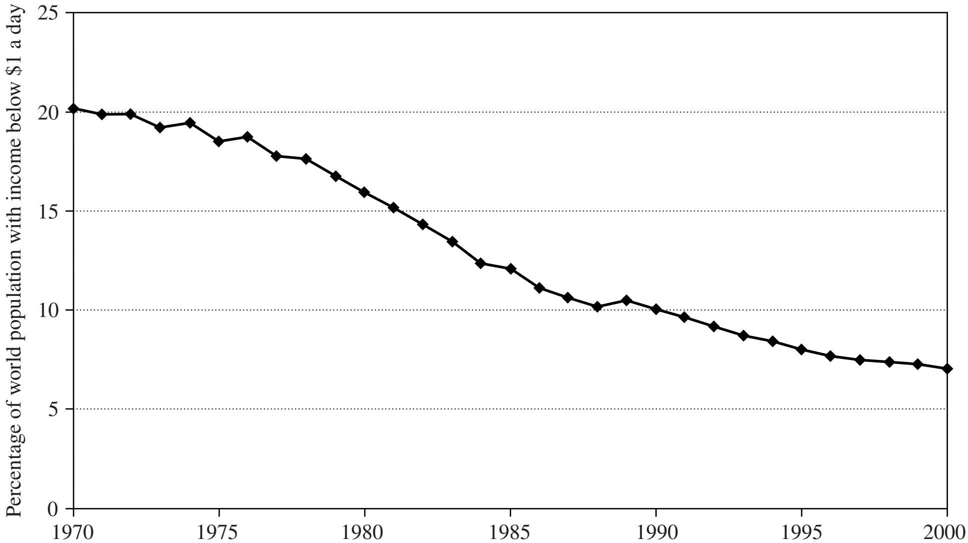

Figure I.7
World poverty rates. The graphs show the fraction of overall population with income below the poverty line. Source: Sala-i-Martin (2003a).

The clear conclusion is that economic growth led to substantial reductions in the world's poverty rates and head counts over the last thirty years. As mentioned earlier, this outcome was not inevitable: if aggregate growth had been accompanied by substantial increases in income inequality, it would have been possible for the mean of the income distribution to increase but also for the fraction of the distribution below a specified poverty threshold to also increase. Sala-i-Martin (2003a) shows that, even though this result is theoretically possible, the world did not behave this way over the last thirty years. Moreover, he also shows that world income inequality actually declined slightly between 1980 and 2000. This conclusion holds whether inequality is measured by the Gini coefficient, the Theil Index, the mean logarithmic deviation, various Atkinson indexes, the variance of log-income, or the coefficient of variation.

Sala-i-Martin (2003a) decomposes the world into regions and notes that poverty erradication has been most pronounced in the regions where growth has been the largest. Figure I.8 reports poverty rates for the poorest regions of the world: East Asia, South Asia, Latin America, Africa, the Middle East and North Africa (MENA), and Eastern Europe and Central Asia. In 1970, three of these regions had poverty rates close to or above 30 percent. Two of them (East Asia and South Asia) have experienced substantial reductions in poverty

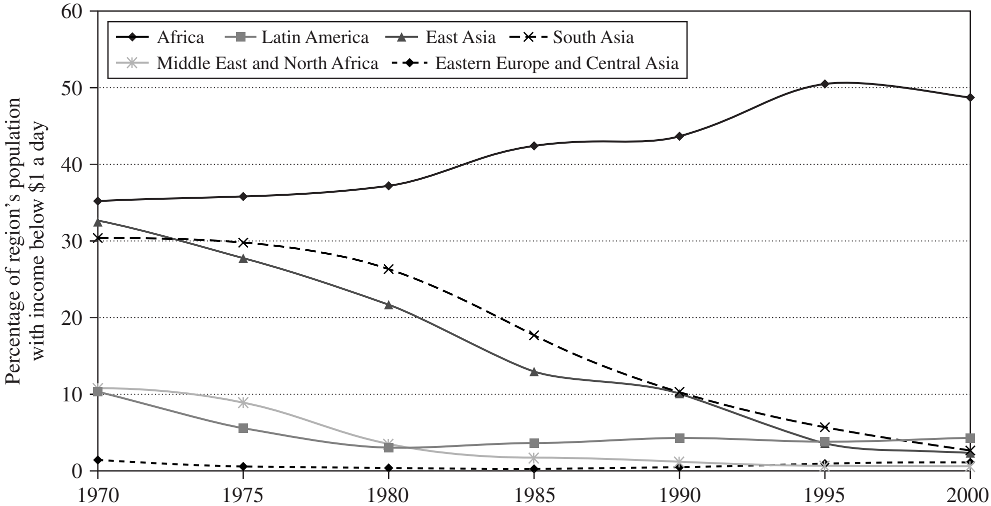

Figure I.8
Regional poverty rates. The graphs show the fraction of each region's population with income below the poverty line. The regions are the ones defined by the World Bank: East Asia, South Asia, Latin America, Africa, the Middle East and North Africa (MENA), and Eastern Europe and Central Asia. Source: Sala-i-Martin (2003a).

rates. These are the regions that also experienced large positive aggregate growth rates. The other region (Africa) has witnessed a dramatic increase in poverty rates over the last thirty years. We also know that per capita growth rates have been negative or close to zero for most countries in Africa. Figure I.8 also shows that two regions had poverty rates near 10 percent in 1970: Latin America and MENA. Both have experienced reductions in poverty rates. Latin America witnessed dramatic gains in the 1970s, when growth rates were substantial, but suffered a setback during the 1980s (the "lost decade," which featured negative growth rates). Poverty rates in Latin America stabilized during the 1990s. Poverty rates in MENA declined slightly between 1970 and 1975. The decline was very large during the high-growth decade that followed the oil shocks and then stabilized when aggregate growth stopped.

Finally, Eastern Europe and Central Asia (a region that includes the former Soviet Union) started off with very small poverty rates. The rates multiplied by a factor of 10 between 1989 and 2000. There are two reasons for the explosion of poverty rates in Eastern Europe and Central Asia. One is the huge increase in inequality that followed the collapse of the communist system. The second factor is the dismal aggregate growth performance of these countries. Notice, however, that the average levels of income for these countries remain far above the levels of Africa or even Asia. Therefore, even after the deterioration in mean income and the rise of income dispersion, poverty rates remain relatively low in Eastern Europe and Central Asia.

I.3 Empirical Regularities about Economic Growth

Kaldor (1963) listed a number of stylized facts that he thought typified the process of economic growth:

1. Per capita output grows over time, and its growth rate does not tend to diminish.
2. Physical capital per worker grows over time.
3. The rate of return to capital is nearly constant.
4. The ratio of physical capital to output is nearly constant.
5. The shares of labor and physical capital in national income are nearly constant.
6. The growth rate of output per worker differs substantially across countries.[8]

Fact 6 accords with the cross-country data that we have already discussed. Facts 1, 2, 4, and 5 seem to fit reasonably well with the long-term data for currently developed countries. For discussions of the stability of the long-run ratio of physical capital to GDP in Japan, Germany, Italy, the United Kingdom, and the United States, see Maddison (1982, chapter 3). For indications of the long-term stability of factor shares in the United States, see Denison (1974, appendix J) and Jorgenson, Gollop, and Fraumeni (1987, table 9.3). Young (1995) reports that factor shares were reasonably stable in four East Asian countries—Hong Kong, Singapore, South Korea, and Taiwan—from the early or middle 1960s through 1990. Studies of seven developed countries—Canada, France, Germany, Italy, Japan, the Netherlands, and the United Kingdom—indicate that factor shares are similar to those in the United States (Christensen, Cummings, and Jorgenson, 1980, and Dougherty, 1991). In some Latin-American countries considered by Elias (1990), the capital shares tend, however, to be higher than those in the United States.

Kaldor's claimed fact 3 on the stability of real rates of return appears to be heavily influenced by the experience of the United Kingdom; in this case, the real interest rate seems

8. Kuznets (1973, 1981) brings out other characteristics of modern economic growth. He notes the rapid rate of structural transformation, which includes shifts from agriculture to industry to services. This process involves urbanization, shifts from home work to employee status, and an increasing role for formal education. He also argues that modern growth involves an increased role for foreign commerce and that technological progress implies reduced reliance on natural resources. Finally, he discusses the growing importance of government: "The spread of modern economic growth placed greater emphasis on the importance and need for organization in national sovereign units. . . . The sovereign state unit was of critical importance as the formulator of the rules under which economic activity was to be carried on; as a referee . . . ; and as provider of infrastructure" (1981, p. 59).

to have no long-run trend (see Barro, 1987, figures 4 and 7). For the United States, however, the long-term data suggest a moderate decline of real interest rates (Barro, 1997, table 11.1). Real rates of return in some fast-growing countries, such as South Korea and Singapore, are much higher than those in the United States but have declined over time (Young, 1995). Thus it seems likely that Kaldor's hypothesis of a roughly stable real rate of return should be replaced by a tendency for returns to fall over some range as an economy develops.

We can use the data presented in chapter 12 to assess the long-run tendencies of the growth rate of real per capita GDP. Tables 12.10 and 12.11 contain figures from Angus Maddison for 31 countries over periods of roughly a century. These numbers basically exhaust the available information about growth over very long time intervals.

Table 12.10 applies to 16 currently developed countries, the major countries in Europe plus the United States, Canada, and Australia. These data show an average per capita growth rate of 1.9 percent per year over roughly a century, with a breakdown by 20-year periods as shown in table I.1. These numbers are consistent with Kaldor's proposition that the growth rate of real per capita GDP has no secular tendency to decline; in fact, the periods following World War II show growth rates well above the long-run average. The reduction in the growth rate from 3.7 percent per year in 1950–70 to 2.2 percent per year in 1970–90 corresponds to the often-discussed *productivity slowdown*. It is apparent from the table, however, that the growth rate for 1970–90 is high in relation to the long-term history.

Table 12.11 contains figures for 15 currently less-developed countries in Asia and Latin America. In this case, the average long-run growth rate from 1900 to 1987 is 1.4 percent per year, and the breakdown into four subperiods is as shown in table I.2. Again, the post–World War II period (here, 1950–87) shows growth rates well above the long-term average.

Table I.1
Long-Term Growth Rates for Currently Developed Countries

Period	Growth Rate (percent per year)	Number of Countries
1830–50	0.9	10
1850–70	1.2	11
1870–90	1.2	13
1890–10	1.5	14
1910–30	1.3	16
1930–50	1.4	16
1950–70	3.7	16
1970–90	2.2	16

Source: Table 12.10.
Note: The growth rates are simple averages for the countries with data.

Table I.2
Long-Term Growth Rates for Currently Less-Developed Countries

Period	Growth Rate (percent per year)	Number of Countries
1900–13	1.2	15
1913–50	0.4	15
1950–73	2.6	15
1973–87	2.4	15

Source: Table 12.11 in chapter 12.
Note: The growth rates are simple averages for the countries with data.

The information depicted in figures I.1–I.3 applies to the behavior of real per capita GDP for over 100 countries from 1960 to 2000. We can use these data to extend the set of stylized facts that was provided by Kaldor. One pattern in the cross-country data is that the growth rate of per capita GDP from 1960 to 2000 is essentially uncorrelated with the level of per capita GDP in 1960 (see chapter 12). In the terminology developed in chapter 1, we shall refer to a tendency for the poor to grow faster than the rich as β convergence. Thus the simple relationship between growth and the starting position for a broad cross section of countries does not reveal β convergence. This kind of convergence does appear if we limit attention to more homogeneous groups of economies, such as the U.S. states, regions of several European countries, and prefectures of Japan (see Barro and Sala-i-Martin, 1991, 1992a, and 1992b, and chapter 11). In these cases, the poorer places tend to grow faster than the richer ones. This behavior also appears in the cross-country data if we limit the sample to a relatively homogeneous collection of currently prosperous places, such as the OECD countries (see Baumol, 1986; DeLong, 1988).

We say in chapter 1 that *conditional* β convergence applies if the growth rate of per capita GDP is negatively related to the starting level of per capita GDP after holding fixed some other variables, such as initial levels of human capital, measures of government policies, the propensities to save and have children, and so on. The broad cross-country sample—that is, the data set that does not show β convergence in an absolute sense—clearly reveals β convergence in this conditional context (see Barro, 1991; Barro and Sala-i-Martin, 1992a; and Mankiw, Romer, and Weil, 1992). The rate of convergence is, however, only about 2 percent per year. Thus, it takes about 35 years for an economy to eliminate one-half of the gap between its initial per capita GDP and its long-run or target level of per capita GDP. (The target tends to grow over time.)

The results in chapter 12 show that a number of variables are significantly related to the growth rate of per capita GDP, once the starting level of per capita GDP is held constant. For example, growth depends positively on the initial quantity of human capital in the form of educational attainment and health, positively on maintenance of the rule of law and the

Table I.3
Ratios to GDP of Gross Domestic Investment and Gross National Saving (percent)

Period	Australia	Canada	France	India	Japan	Korea	United Kingdom	United States
				1. Gross Domestic Investment				
1870–89	16.5	16.0	12.8	—	—	—	9.3	19.8
1890–09	13.7	17.2	14.0	—	14.0	—	9.4	17.9
1910–29	17.4	19.8	—	6.4	16.6	5.1[a]	6.7	17.2
1930–49	13.3	13.1	—	8.4	20.5	—	8.1	12.7
1950–69	26.3	23.8	22.6	14.0	31.8	16.3[b]	17.2	18.9
1970–89	24.9	22.8	23.2	20.2	31.9	29.1	18.2	18.7
				2. Gross National Saving				
1870–89	11.2	9.1	12.8	—	—	—	13.9	19.1
1890–09	12.2	11.5	14.9	—	12.0	—	13.1	18.4
1910–29	13.6	16.0	—	6.4	17.1	2.38	9.6	18.9
1930–49	13.0	15.6	—	7.7	19.8	—	4.8	14.1
1950–69	24.0	22.3	22.8	12.2	32.1	5.9[b]	17.7	19.6
1970–89	22.9	22.1	23.4	19.4	33.7	26.2	19.4	18.5

Source: Maddison (1992).
[a] 1911–29
[b] 1951–69

ratio of investment to GDP, and negatively on fertility rates and the ratio of government consumption spending to GDP.

We can assess regularities in investment and saving ratios by using the long-term data in Maddison (1992). He provides long-term information for a few countries on the ratios of gross domestic investment to GDP and of gross national saving (the sum of domestic and net foreign investment) to GDP. Averages of the investment and saving ratios over 20-year intervals for the eight countries that have enough data for a long-period analysis are shown in table I.3. For an individual country, the table indicates that the time paths of domestic investment and national saving are usually similar. Domestic investment was, however, substantially higher than national saving (that is, borrowing from abroad was large) for Australia and Canada from 1870 to 1929, for Japan from 1890 to 1909, for the United Kingdom from 1930 to 1949, and for Korea from 1950 to 1969 (in fact, through the early 1980s). National saving was much higher than domestic investment (lending abroad was substantial) for the United Kingdom from 1870 to 1929 and for the United States from 1930 to 1949.

For the United States, the striking observation from the table is the stability over time of the ratios for domestic investment and national saving. The only exception is the relatively low values from 1930 to 1949, the period of the Great Depression and World War II. The United States is, however, an outlier with respect to the stability of its investment and saving ratios; the data for the other seven countries show a clear increase in these ratios over time. In particular, the ratios for 1950–89 are, in all cases, substantially greater than those from before World War II. The long-term data therefore suggest that the ratios to GDP of gross domestic investment and gross national saving tend to rise as an economy develops, at least over some range. The assumption of a constant gross saving ratio, which appears in chapter 1 in the Solow–Swan model, misses this regularity in the data.

The cross-country data also reveal some regularities with respect to fertility rates and, hence, rates of population growth. For most countries, the fertility rate tends to decline with increases in per capita GDP. For the poorest countries, however, the fertility rate may rise with per capita GDP, as Malthus (1798) predicted. Even stronger relations exist between educational attainment and fertility. Except for the most advanced countries, female schooling is negatively related with the fertility rate, whereas male schooling is positively related with the fertility rate. The net effect of these forces is that the fertility rate—and the rate of population growth—tend to fall over some range as an economy develops. The assumption of an exogenous, constant rate of population growth—another element of the Solow–Swan model—conflicts with this empirical pattern.

I.4 A Brief History of Modern Growth Theory

Classical economists, such as Adam Smith (1776), David Ricardo (1817), and Thomas Malthus (1798), and, much later, Frank Ramsey (1928), Allyn Young (1928), Frank Knight (1944), and Joseph Schumpeter (1934), provided many of the basic ingredients that appear in modern theories of economic growth. These ideas include the basic approaches of competitive behavior and equilibrium dynamics, the role of diminishing returns and its relation to the accumulation of physical and human capital, the interplay between per capita income and the growth rate of population, the effects of technological progress in the forms of increased specialization of labor and discoveries of new goods and methods of production, and the role of monopoly power as an incentive for technological advance.

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 the 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 per capita GDP, relative to the long-run or steady-state position, the faster 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 an analysis of imperfect competition.[9] 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.[10] 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.

After the mid-1980s, research on economic growth experienced a 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 was only a first step; to go further, one had to escape the straitjacket of the neoclassical growth model, in which the long-term per capita growth rate was 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

9. 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).

10. However, recent research has shown how to extend the neoclassical growth model to allow for heterogeneity among households (Caselli and Ventura, 2000) and to incorporate time-inconsistent preferences (Barro, 1999).

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 to 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 included 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, 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. This research program remained active through the 1990s and has been applied, for example, to understanding scale effects in the growth process (Jones, 1999), analyzing whether technological progress will be labor or capital augmenting (Acemoglu, 2002), and assessing the role of competition in the growth process (Aghion et al., 2001, 2002).

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. Some recent empirical work has verified the importance of technological diffusion in the convergence process.

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 into 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 1990s is that the recent research pays close attention to empirical implications and to the relation between theory and data. However, much of this applied perspective involved applications of empirical hypotheses from the older theory, notably the neoclassical growth model's prediction of conditional convergence. The cross-country regressions motivated by the neoclassical model surely became a fixture of research in the 1990s. An interesting recent development in this area, which we explore in chapter 12, involves assessment of the robustness of these kinds of estimates. Other empirical 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.

I.5 Some Highlights of the Second Edition

This second edition of *Economic Growth* includes changes throughout the book. We mention here a few of the highlights. In this introduction we already described new estimates of the distribution of income of individuals throughout the world from 1970 to 2000.

Chapter 1 has been made easier and more accessible. We added a section on markets in the Solow–Swan model. We also discussed the nature of the theoretical dissatisfaction with neoclassical theory that led to the emergence of endogenous growth models with imperfect competition.

Chapter 2 expands the treatment of the basic neoclassical growth model to allow for heterogeneity of households. There is an improved approach to ruling out "undersaving" paths and for deriving and using transversality conditions. We also include an analysis of models with nonconstant time-preference rates.

Chapter 3 has various extensions to the basic neoclassical growth model, including an expanded treatment of the government sector. The framework allows for various forms of tax rates and allows for a clear distinction between taxes on capital income and taxes on labor or consumption.

Chapters 6 and 7 discuss models of endogenous technological progress. The new material includes an analysis of the role and source of scale effects in these models. We refer in chapter 6 to Thomas Jefferson's mostly negative views on patents as a mechanism for motivating inventions. Chapter 7 has an improved analysis of models where technological advances take the form of quality improvements. We have particularly improved the treatment of the interplay between industry leaders and outsiders and, hence, of the role of outside competition in the growth process.

Chapter 8 has a model of technological diffusion. The basic model is improved, and the theoretical results are related to recent empirical findings.

Chapter 9 has an extended treatment of endogenous population growth. Chapter 10 has an improved analysis of growth accounting, including its relation to theories of endogenous technological progress. Chapter 11, which deals with regional data sets, extends the analysis of U.S. states through 2000.

In chapter 12 we include an updated treatment of cross-country growth regressions, using the new Summers–Heston data set, Penn World Tables version 6.1, which has data through 2000 (see Heston, Summers, and Aten, 2002). We also discuss in this chapter various issues about the reliability of estimates from cross-country regressions, including ways to assess the robustness of the results.

1 Growth Models with Exogenous Saving Rates (the Solow–Swan Model)

1.1 The Basic Structure

The first question we ask in this chapter is whether it is possible for an economy to enjoy positive growth rates forever by simply saving and investing in its capital stock. A look at the cross-country data from 1960 to 2000 shows that the average annual growth rate of real per capita GDP for 112 countries was 1.8 percent, and the average ratio of gross investment to GDP was 16 percent.[1] However, for 38 sub-Saharan African countries, the average growth rate was only 0.6 percent, and the average investment ratio was only 10 percent. At the other end, for nine East Asian "miracle" economies, the average growth rate was 4.9 percent, and the average investment ratio was 25 percent. These observations suggest that growth and investment rates are positively related. However, before we get too excited with this relationship, we might note that, for 23 OECD countries, the average growth rate was 2.7 percent—lower than that for the East Asian miracles—whereas the average investment ratio was 24 percent—about the same as that for East Asia. Thus, although investment propensities cannot be the whole story, it makes sense as a starting point to try to relate the growth rate of an economy to its willingness to save and invest. To this end, it will be useful to begin with a simple model in which the only possible source of per capita growth is the accumulation of physical capital.

Most of the growth models 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 them to produce goods that they sell to households or other firms. Firms have access to a technology 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.

Although this general structure applies to most growth models, it is convenient to start our analysis by using 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. In the real world, production takes place using many different inputs to production. We summarize all of them into just three: physical capital $K(t)$, labor $L(t)$, and knowledge $T(t)$. The production

1. These data—from Penn World Tables version 6.1—are described in Summers and Heston (1991) and Heston, Summers, and Aten (2002). We discuss these data in chapter 12.

function takes the form

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

where $Y(t)$ is the flow of output produced at time t.

Capital, $K(t)$, represents the durable physical inputs, such as machines, buildings, pencils, and so on. These goods were produced sometime in the past by a production function of the form of equation (1.1). It is important to notice that these inputs cannot be used by multiple producers simultaneously. This last characteristic is known as *rivalry*—a good is *rival* if it cannot be used by several users at the same time.

The second input to the production function is labor, $L(t)$, and it represents the inputs associated with the human body. This input includes the number of workers and the amount of time they work, as well as their physical strength, skills, and health. Labor is also a *rival* input, because a worker cannot work on one activity without reducing the time available for other activities.

The third input is the level of knowledge or technology, $T(t)$. Workers and machines cannot produce anything without a *formula* or *blueprint* that shows them how to do it. This blueprint is what we call *knowledge or technology*. Technology can improve over time—for example, the same amount of capital and labor yields a larger quantity of output in 2000 than in 1900 because the technology employed in 2000 is superior. Technology can also differ across countries—for example, the same amount of capital and labor yields a larger quantity of output in Japan than in Zambia because the technology available in Japan is better. The important distinctive characteristic of knowledge is that it is a *nonrival good:* two or more producers can use the same formula at the same time.[2] Hence, two producers that each want to produce Y units of output will each have to use a different set of machines and workers, but they can use the same formula. This property of nonrivalry turns out to have important implications for the interactions between technology and economic growth.[3]

2. The concepts of *nonrivalry* and *public good* are often confused in the literature. *Public goods* are *nonrival* (they can be used by many people simultaneously) and also *nonexcludable* (it is technologically or legally impossible to prevent people from using such goods). The key characteristic of knowledge is nonrivalry. Some formulas or blueprints are nonexcludable (for example, calculus formulas on which there are no property rights), whereas others are excludable (for example, the formulas used to produce pharmaceutical products while they are protected by patents). These properties of ideas were well understood by Thomas Jefferson, who said in a letter of August 13, 1813, to Isaac McPherson: "If nature has made any one thing less susceptible than all others of exclusive property, it is the actions of the thinking power called an idea, which an individual may exclusively possess as long as he keeps it to himself; but the moment it is divulged, it forces itself into the possession of everyone, and the receiver cannot dispossess himself of it. Its peculiar character, too, is that no one possesses the less, because every other possesses the whole of it. He who receives an idea from me, receives instruction himself without lessening mine" (available on the Internet from the Thomas Jefferson Papers at the Library of Congress, lcweb2.loc.gov/ammem/mtjhtml/mtjhome.html).

3. Government policies, which depend on laws and institutions, would also affect the output of an economy. Since basic public institutions are nonrival, we can include these factors in $T(t)$ in the production function.

We assume a one-sector production technology in which output is a homogeneous good that can be consumed, $C(t)$, or invested, $I(t)$. Investment is used to create new units of physical capital, $K(t)$, or to replace old, depreciated capital. 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.

In this chapter we imagine 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.) We also start with the assumption that there are no government purchases of goods and services. (Chapter 4 deals with government purchases.) In a closed economy with no public spending, all output is devoted to consumption or gross investment,[4] so $Y(t) = C(t) + I(t)$. By subtracting $C(t)$ from both sides and realizing that output equals income, we get that, in this simple economy, the amount saved, $S(t) \equiv Y(t) - C(t)$, equals the amount invested, $I(t)$.

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 of the state of the economy, a 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, $0 \leq s(\cdot) = s \leq 1$. We use this constant-saving-rate specification in this chapter because it brings out a large number of results in a clear way. Given that saving must equal investment, $S(t) = I(t)$, it follows that the *saving rate* equals the *investment rate*. In other words, the saving rate of a closed economy represents the fraction of GDP that an economy devotes to investment.

We assume that capital is a homogeneous good that 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. Before evaporating, however, all units of capital are assumed to be equally productive, regardless of when they were originally produced.

4. In an open economy with government spending, the condition is

$$Y(t) - r \cdot D(t) = C(t) + I(t) + G(t) + NX(t)$$

where $D(t)$ is international debt, r is the international real interest rate, $G(t)$ is public spending, and $NX(t)$ is net exports. In this chapter we assume that there is no public spending, so that $G(t) = 0$, and that the economy is closed, so that $D(t) = NX(t) = 0$.

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

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

where a dot over a variable, such as $\dot{K}(t)$, denotes differentiation with respect to time, $\dot{K}(t) \equiv \partial K(t)/\partial t$ (a convention that we use throughout the book) and $0 \leq s \leq 1$. Equation (1.2) determines the dynamics of K for a given technology and labor.

The labor input, L, varies over time because of population growth, changes in participation rates, shifts in the amount of time worked by the typical worker, and improvements in the skills and quality of workers. In this chapter, we simplify by assuming that everybody works the same amount of time and that everyone has the same constant skill, which we normalize to one. Thus we identify the labor input with the total population. We analyze the accumulation of skills or human capital in chapter 5 and the choice between labor and leisure in chapter 9.

The growth of population reflects the behavior of fertility, mortality, and migration, which we study in chapter 9. In this chapter, we simplify by assuming that population grows at a constant, exogenous rate, $\dot{L}/L = n \geq 0$, without using any resources. 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

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

To highlight the role of capital accumulation, we start with the assumption that the level of technology, $T(t)$, is a constant. This assumption will be relaxed later.

If $L(t)$ is given from equation (1.3) and technological progress is absent, then equation (1.2) determines the time paths of capital, $K(t)$, and output, $Y(t)$. Once we know how capital or GDP changes over time, the growth rates of these variables are also determined. In the next sections, we show that this behavior depends crucially on the properties of the production function, $F(\cdot)$.

1.2 The Neoclassical Model of Solow and Swan

1.2.1 The Neoclassical Production Function

The process of economic growth depends on the shape of the production function. We initially consider the neoclassical production function. We say that a production function, $F(K, L, T)$, is *neoclassical* if the following properties are satisfied:[5]

5. We ignore time subscripts to simplify notation.

1. Constant returns to scale. The function $F(\cdot)$ exhibits constant returns to scale. That is, if we multiply capital and labor by the same positive constant, λ, we get λ the amount of output:

$$F(\lambda K, \lambda L, T) = \lambda \cdot F(K, L, T) \quad \text{for all } \lambda > 0 \tag{1.4}$$

This property is also known as *homogeneity of degree one in K and L.* It is important to note that the definition of scale includes only the two rival inputs, capital and labor. In other words, we did not define constant returns to scale as $F(\lambda K, \lambda L, \lambda T) = \lambda \cdot F(K, L, T)$.

To get some intuition on why our assumption makes economic sense, we can use the following *replication argument.* Imagine that plant 1 produces Y units of output using the production function F and combining K and L units of capital and labor, respectively, and using formula T. It makes sense to assume that if we create an identical plant somewhere else (that is, if we *replicate* the plant), we should be able to produce the same amount of output. In order to replicate the plant, however, we need a new set of machines and workers, but we can use the same formula in both plants. The reason is that, while capital and labor are rival goods, the formula is a nonrival good and can be used in both plants at the same time. Hence, because technology is a nonrival input, our definition of returns to scale makes sense.

2. Positive and diminishing returns to private inputs. For all $K > 0$ and $L > 0$, $F(\cdot)$ exhibits positive and diminishing marginal products with respect to each input:

$$\begin{aligned} \frac{\partial F}{\partial K} > 0, \qquad \frac{\partial^2 F}{\partial K^2} < 0 \\ \frac{\partial F}{\partial L} > 0, \qquad \frac{\partial^2 F}{\partial L^2} < 0 \end{aligned} \tag{1.5}$$

Thus, the neoclassical technology assumes that, holding constant the levels of technology and labor, each additional unit of capital delivers positive additions to output, but these additions decrease as the number of machines rises. The same property is assumed for labor.

3. Inada conditions. The third defining characteristic of the neoclassical production function is that 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:

$$\begin{aligned} \lim_{K \to 0} \left(\frac{\partial F}{\partial K} \right) = \lim_{L \to 0} \left(\frac{\partial F}{\partial L} \right) = \infty \\ \lim_{K \to \infty} \left(\frac{\partial F}{\partial K} \right) = \lim_{L \to \infty} \left(\frac{\partial F}{\partial L} \right) = 0 \end{aligned} \tag{1.6}$$

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

4. Essentiality. Some economists add the assumption of *essentiality* to the definition of a neoclassical production function. An input is essential if a strictly positive amount is needed to produce a positive amount of output. We show in the appendix that the three neoclassical properties in equations (1.4)–(1.6) imply that each input is *essential* for production, that is, $F(0, L) = F(K, 0) = 0$. The three properties of the neoclassical production function also imply that output goes to infinity as either input goes to infinity, another property that is proven in the appendix.

Per Capita Variables When we say that a country is rich or poor, we tend to think in terms of output or consumption per person. In other words, we do not think that India is richer than the Netherlands, even though India produces a lot more GDP, because, once we divide by the number of citizens, the amount of income each person gets on average is a lot smaller in India than in the Netherlands. To capture this property, we construct the model in per capita terms and study primarily the dynamic behavior of the per capita quantities of GDP, consumption, and capital.

Since the definition of constant returns to scale applies to all values of λ, it also applies to $\lambda = 1/L$. Hence, output can be written as

$$Y = F(K, L, T) = L \cdot F(K/L, 1, T) = L \cdot f(k) \tag{1.7}$$

where $k \equiv K/L$ is capital per worker, $y \equiv Y/L$ is output per worker, and the function $f(k)$ is defined to equal $F(k, 1, T)$.[6] This result means that the production function can be expressed in *intensive form* (that is, in *per worker* or *per capita* form) as

$$y = f(k) \tag{1.8}$$

In other words, the production function exhibits no "scale effects": production per person is determined by the amount of physical capital each person has access to and, holding constant k, having more or fewer workers does not affect total output per person. Consequently, very large economies, such as China or India, can have less output or income per person than very small economies, such as Switzerland or the Netherlands.

We can differentiate this condition $Y = L \cdot f(k)$ 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

$$\partial Y/\partial K = f'(k) \tag{1.9}$$

$$\partial Y/\partial L = f(k) - k \cdot f'(k) \tag{1.10}$$

6. Since T is assumed to be constant, it is one of the parameters implicit in the definition of $f(k)$.

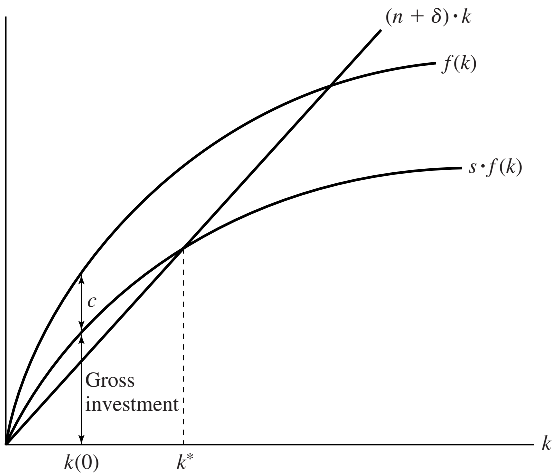

Figure 1.1
The Solow–Swan model. The curve for gross investment, $s \cdot f(k)$, is proportional to the production function, $f(k)$. Consumption per person equals the vertical distance between $f(k)$ and $s \cdot f(k)$. Effective depreciation (for k) is given by $(n + \delta) \cdot k$, a straight line from the origin. The change in k is given by the vertical distance between $s \cdot f(k)$ and $(n + \delta) \cdot k$. The steady-state level of capital, k^*, is determined at the intersection of the $s \cdot f(k)$ curve with the $(n + \delta) \cdot k$ line.

The Inada conditions imply $\lim_{k\to 0}[f'(k)] = \infty$ and $\lim_{k\to\infty}[f'(k)] = 0$. Figure 1.1 shows the neoclassical production in per capita terms: it goes through zero; it is vertical at zero, upward sloping, and concave; and its slope asymptotes to zero as k goes to infinity.

A Cobb–Douglas Example One simple production function that is often thought to provide a reasonable description of actual economies is the Cobb–Douglas function,[7]

$$Y = AK^{\alpha}L^{1-\alpha} \tag{1.11}$$

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

$$y = Ak^{\alpha} \tag{1.12}$$

7. Douglas is Paul H. Douglas, who was a labor economist at the University of Chicago and later a U.S. Senator from Illinois. Cobb is Charles W. Cobb, who was a mathematician at Amherst. Douglas (1972, pp. 46–47) says that he consulted with Cobb in 1927 on how to come up with a production function that fit his empirical equations for production, employment, and capital stock in U.S. manufacturing. Interestingly, Douglas says that the functional form was developed earlier by Philip Wicksteed, thus providing another example of Stigler's Law (whereby nothing is named after the person who invented it).

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.

The key property of the Cobb–Douglas production function is the behavior of factor income shares. In a competitive economy, as discussed in section 1.2.3, capital and labor are each paid their marginal products; that is, the marginal product of capital equals the rental price R, and the marginal product of labor equals the wage rate w. Hence, each unit of capital is paid $R = f'(k) = \alpha A k^{\alpha-1}$, and each unit of labor is paid $w = f(k) - k \cdot f'(k) = (1-\alpha) \cdot Ak^{\alpha}$. The capital share of income is then $Rk/f(k) = \alpha$, and the labor share is $w/f(k) = 1 - a$. Thus, in a competitive setting, the factor income shares are constant—independent of k—when the production function is Cobb–Douglas.

1.2.2 The Fundamental Equation of the Solow–Swan Model

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 equation (1.2). If we divide both sides of this equation by L, 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. Hence, it is not an ordinary differential equation that can be easily solved. In order to transform it into a differential equation in terms of k, we can take the derivative of $k \equiv K/L$ with respect to time to get

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

where $n = \dot{L}/L$. If we substitute this result into the expression for $\dot{K}/L$, we can rearrange terms to get

$$\dot{k} = s \cdot f(k) - (n + \delta) \cdot k \tag{1.13}$$

Equation (1.13) is the fundamental differential equation of the Solow–Swan model. This nonlinear equation depends only on k.

The term $n + \delta$ on the right-hand side of equation (1.13) can be thought of as the effective depreciation rate for the capital-labor ratio, $k \equiv K/L$. If the saving rate, s, were 0, capital per person would decline partly due to depreciation of capital at the rate δ and partly due to the increase in the number of persons at the rate n.

Figure 1.1 shows the workings of equation (1.13). The upper curve is the production function, $f(k)$. The term $(n+\delta)\cdot k$, which appears in equation (1.13), is drawn in figure 1.1 as a straight line from the origin with the positive slope $n+\delta$. The term $s\cdot f(k)$ in equation (1.13) looks like the production function except for the multiplication by the positive fraction s. Note from the figure that the $s\cdot f(k)$ curve starts from the origin [because $f(0)=0$], has a positive slope [because $f'(k)>0$], and gets flatter as k rises [because $f''(k)<0$]. The Inada conditions imply that the $s\cdot f(k)$ curve is vertical at $k=0$ and becomes flat as k goes to infinity. These properties imply that, other than the origin, the curve $s\cdot f(k)$ and the line $(n+\delta)\cdot k$ cross once and only once.

Consider an economy with the initial capital stock per person $k(0)>0$. Figure 1.1 shows that gross investment per person equals the height of the $s\cdot f(k)$ curve at this point. Consumption per person equals the vertical difference at this point between the $f(k)$ and $s\cdot f(k)$ curves.

1.2.3 Markets

In this section we show that the fundamental equation of the Solow–Swan model can be derived in a framework that explicitly incorporates markets. Instead of owning the technology and keeping the output produced with it, we assume that households own financial assets and labor. Assets deliver a rate of return $r(t)$, and labor is paid the wage rate $w(t)$. The total income received by households is, therefore, the sum of asset and labor income, $r(t)\cdot(\text{assets})+w(t)\cdot L(t)$. Households use the income that they do not consume to accumulate more assets

$$d(\text{assets})/dt=[r\cdot(\text{assets})+w\cdot L]-C \tag{1.14}$$

where, again, time subscripts have been omitted to simplify notation. Divide both sides of equation (1.14) by L, define assets per person as a, and take the derivative of a with respect to time, $\dot{a}=(1/L)\cdot d(\text{assets})/dt-na$, to get that the change in assets per person is given by

$$\dot{a}=(r\cdot a+w)-c-na \tag{1.15}$$

Firms hire labor and capital and use these two inputs with the production technology in equation (1.1) to produce output, which they sell at unit price. We think of firms as renting the services of capital from the households that own it. (None of the results would change if the firms owned the capital, and the households owned shares of stock in the firms.) Hence, the firms' costs of capital are the rental payments, which are proportional to K. This specification assumes that capital services can be increased or decreased without incurring any additional expenses, such as costs for installing machines.

Let R be the rental price for a unit of capital services, and assume again that capital stocks depreciate at the constant rate $\delta \geq 0$. The net rate of return to a household that owns a unit of capital is then $R - \delta$. Households also receive the interest rate r on funds lent to other households. In the absence of uncertainty, capital and loans are perfect substitutes as stores of value and, as a result, they must deliver the same return, so $r = R - \delta$ or, equivalently, $R = r + \delta$.

The representative firm's flow of net receipts or profit at any point in time is given by

$$\pi = F(K, L, T) - (r + \delta) \cdot K - wL \tag{1.16}$$

that is, gross receipts from the sale of output, $F(K, L, T)$, less the factor payments, which are rentals to capital, $(r + \delta) \cdot K$, and wages to workers, wL. Technology is assumed to be available for free, so no payment is needed to rent the formula used in the process of production. We assume that the firm seeks to maximize the present value of profits. Because the firm rents capital and labor services and has no adjustment costs, there are no intertemporal elements in the firm's maximization problem.[8] (The problem becomes intertemporal when we introduce adjustment costs for capital in chapter 3.)

Consider a firm of arbitrary scale, say with level of labor input L. Because the production function exhibits constant returns to scale, the profit for this firm, which is given by equation (1.16), can be written as

$$\pi = L \cdot [f(k) - (r + \delta) \cdot k - w] \tag{1.17}$$

A competitive firm, which takes r and w as given, maximizes profit for given L by setting

$$f'(k) = r + \delta \tag{1.18}$$

That is, the firm chooses the ratio of capital to labor to equate the marginal product of capital to the rental price.

The resulting level of profit is positive, zero, or negative depending on the value of w. If profit is positive, the firm could attain infinite profits by choosing an infinite scale. If profit is negative, the firm would contract its scale to zero. Therefore, in a full market equilibrium, w must be such that profit equals zero; that is, the total of the factor payments, $(r + \delta) \cdot K + wL$, equals the gross receipts in equation (1.17). In this case, the firm is indifferent about its scale.

8. In chapter 2 we show that dynamic firms would maximize the present discounted value of all future profits, which is given if r is constant by $\int_0^\infty L \cdot [f(k) - (r + \delta) \cdot k - w] \cdot e^{-rt} dt$. Because the problem does not involve any dynamic constraint, the firm maximizes static profits at all points in time. In fact, this dynamic problem is nothing but a sequence of static problems.

For profit to be zero, the wage rate has to equal the marginal product of labor corresponding to the value of k that satisfies equation (1.18):

$$[f(k) - k \cdot f'(k)] = w \tag{1.19}$$

It can be readily verified from substitution of equations (1.18) and (1.19) into equation (1.17) that the resulting level of profit equals zero for any value of L. Equivalently, if the factor prices equal the respective marginal products, the factor payments just exhaust the total output (a result that corresponds in mathematics to Euler's theorem).[9]

The model does not determine the scale of an individual, competitive firm that operates with a constant-returns-to-scale production function. The model will, however, determine the capital/labor ratio k, as well as the aggregate level of production, because the aggregate labor force is determined by equation (1.3).

The next step is to define the equilibrium of the economy. In a closed economy, the only asset in positive net supply is capital, because all the borrowing and lending must cancel within the economy. Hence, equilibrium in the asset market requires $a = k$. If we substitute this equality, as well as $r = f'(k) - \delta$ and $w = f(k) - k \cdot f'(k)$, into equation (1.15), we get

$$\dot{k} = f(k) - c - (n + \delta) \cdot k$$

Finally, if we follow Solow–Swan in making the assumption that households consume a constant fraction of their gross income, $c = (1 - s) \cdot f(k)$, we get

$$\dot{k} = s \cdot f(k) - (n + \delta) \cdot k$$

which is the same fundamental equation of the Solow–Swan model that we got in equation (1.13). Hence, introducing competitive markets into the Solow–Swan model does not change any of the main results.[10]

1.2.4 The Steady State

We now have the necessary tools to analyze the behavior of the model over time. We first consider the *long run* or *steady state,* and then we describe the *short run* or *transitional dynamics*. We define a *steady state* as a situation in which the various quantities grow at

9. Euler's theorem says that if a function $F(K, L)$ is homogeneous of degree one in K and L, then $F(K, L) = F_K \cdot K + F_L \cdot L$. This result can be proven using the equations $F(K, L) = L \cdot f(k)$, $F_K = f'(k)$, and $F_L = f(k) - k \cdot f'(k)$.

10. Note that, in the previous section and here, we assumed that each person saved a constant fraction of his or her gross income. We could have assumed instead that each person saved a constant fraction of his or her net income, $f(k) - \delta k$, which in the market setup equals $ra + w$. In this case, the fundamental equation of the Solow–Swan model would be $\dot{k} = s \cdot f(k) - (s\delta + n) \cdot k$. Again, the same equation applies to the household-producer and market setups.

constant (perhaps zero) rates.[11] In the Solow–Swan model, the steady state corresponds to $\dot{k} = 0$ in equation (1.13),[12] that is, to an intersection of the $s \cdot f(k)$ curve with the $(n+\delta) \cdot k$ line in figure 1.1.[13] The corresponding value of k is denoted k^*. (We focus here on the intersection at $k > 0$ and neglect the one at $k = 0$.) Algebraically, k^* satisfies the condition

$$s \cdot f(k^*) = (n+\delta) \cdot k^* \tag{1.20}$$

Since k is constant in the steady state, y and c are also constant at the values $y^* = f(k^*)$ and $c^* = (1-s) \cdot f(k^*)$, respectively. Hence, in the neoclassical model, the per capita quantities k, y, and c do not grow in the steady state. The constancy of the per capita magnitudes means that the levels of variables—K, Y, and C—grow in the steady state at the rate of population growth, n.

Once-and-for-all changes in the level of the technology will be represented by shifts of the production function, $f(\cdot)$. Shifts in the production function, in the saving rate s, in the rate of population growth n, and in the depreciation rate δ, all have effects on the per capita *levels* of the various quantities in the steady state. In figure 1.1, for example, a proportional upward shift of the production function or an increase in s shifts the $s \cdot f(k)$ curve upward and leads thereby to an increase in k^*. An increase in n or δ moves the $(n+\delta) \cdot k$ line upward and leads to a decrease in k^*.

It is important to note that a one-time change in the level of technology, the saving rate, the rate of population growth, and the depreciation rate do not affect the steady-state growth rates of per capita output, capital, and consumption, which are all still equal to zero. For this reason, the model as presently specified will not provide explanations of the determinants of long-run per capita growth.

1.2.5 The Golden Rule of Capital Accumulation and Dynamic Inefficiency

For a given level of A and given values of n and δ, there is a unique steady-state value $k^* > 0$ for each value of the saving rate s. Denote this relation by $k^*(s)$, with $dk^*(s)/ds > 0$. The steady-state level of per capita consumption is $c^* = (1-s) \cdot f[k^*(s)]$. We know from

11. Some economists use the expression *balanced growth path* to describe the state in which all variables grow at a constant rate and use *steady state* to describe the particular case when the growth rate is zero.

12. We can show that k must be constant in the steady state. Divide both sides of equation (1.13) by k to get $\dot{k}/k = s \cdot f(k)/k - (n+\delta)$. The left-hand side is constant, by definition, in the steady state. Since s, n, and δ are all constants, it follows that $f(k)/k$ must be constant in the steady state. The time derivative of $f(k)/k$ equals $-\{[f(k) - kf'(k)]/k\} \cdot (\dot{k}/k)$. The expression $f(k) - kf'(k)$ equals the marginal product of labor (as shown by equation [1.19]) and is positive. Therefore, as long as k is finite, $\dot{k}/k$ must equal 0 in the steady state.

13. The intersection in the range of positive k exists and is unique because $f(0) = 0$, $n+\delta < \lim_{k\to 0}[s \cdot f'(k)] = \infty$, $n+\delta > \lim_{k\to\infty}[s \cdot f'(k)] = 0$, and $f''(k) < 0$.

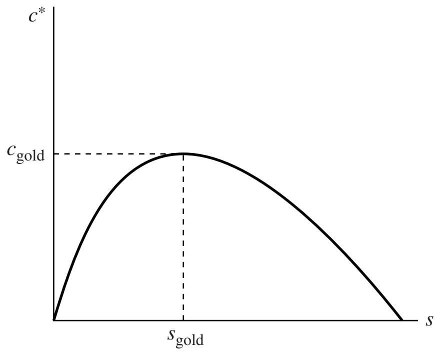

Figure 1.2
The golden rule of capital accumulation. The vertical axis shows the steady-state level of consumption per person that corresponds to each saving rate. The saving rate that maximizes steady-state consumption per person is called the golden-rule saving rate and is denoted by s_{Gold}.

equation (1.20) that $s \cdot f(k^*) = (n + \delta) \cdot k^*$; hence, we can write an expression for c^* as

$$c^*(s) = f[k^*(s)] - (n + \delta) \cdot k^*(s) \tag{1.21}$$

Figure 1.2 shows the relation between c^* and s that is implied by equation (1.21). The quantity c^* is increasing in s for low levels of s and decreasing in s for high values of s. The quantity c^* attains its maximum when the derivative vanishes, that is, when $[f'(k^*) - (n + \delta)] \cdot dk^*/ds = 0$. Since $dk^*/ds > 0$, the term in brackets must equal 0. If we denote the value of k^* that corresponds to the maximum of c^* by k_{gold}, then the condition that determines k_{gold} is

$$f'(k_{gold}) = n + \delta \tag{1.22}$$

The corresponding saving rate can be denoted as s_{gold}, and the associated level of steady-state per capita consumption is given by $c_{gold} = f(k_{gold}) - (n + \delta) \cdot k_{gold}$.

The condition in equation (1.22) is called the *golden rule of capital accumulation* (see Phelps, 1966). The source of this name is the biblical Golden Rule, which states, “Do unto others as you would have others do unto you.” In economic terms, the golden-rule result can be interpreted as “If we provide the same amount of consumption to members of each current and future generation—that is, if we do not provide less to future generations than to ourselves—then the maximum amount of per capita consumption is c_{gold}.”

Figure 1.3 illustrates the workings of the golden rule. The figure considers three possible saving rates, s_1, s_{gold}, and s_2, where $s_1 < s_{gold} < s_2$. Consumption per person, c, in each case equals the vertical distance between the production function, $f(k)$, and the appropriate

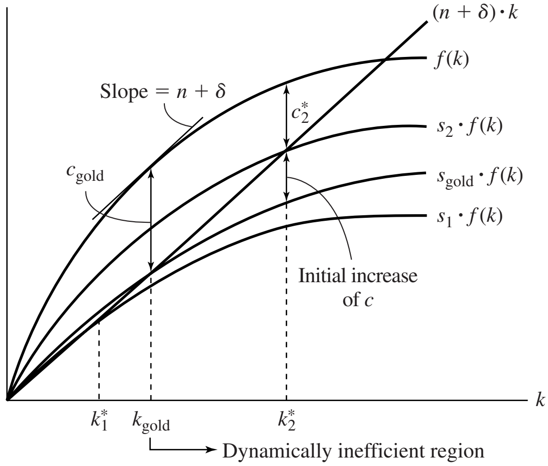

Figure 1.3
The golden rule and dynamic inefficiency. If the saving rate is above the golden rule ($s_2 > s_{\text{gold}}$ in the figure), a reduction in s increases steady-state consumption per person and also raises consumption per person along the transition. Since c increases at all points in time, a saving rate above the golden rule is dynamically inefficient. If the saving rate is below the golden rule ($s_1 < s_{\text{gold}}$ in the figure), an increase in s increases steady-state consumption per person but lowers consumption per person along the transition. The desirability of such a change depends on how households trade off current consumption against future consumption.

$s \cdot f(k)$ curve. For each s, the steady-state value k^* corresponds to the intersection between the $s \cdot f(k)$ curve and the $(n + \delta) \cdot k$ line. The steady-state per capita consumption, c^*, is maximized when $k^* = k_{\text{gold}}$ because the tangent to the production function at this point parallels the $(n + \delta) \cdot k$ line. The saving rate that yields $k^* = k_{\text{gold}}$ is the one that makes the $s \cdot f(k)$ curve cross the $(n + \delta) \cdot k$ line at the value k_{gold}. Since $s_1 < s_{\text{gold}} < s_2$, we also see in the figure that $k_1^* < k_{\text{gold}} < k_2^*$.

An important question is whether some saving rates are better than others. We will be unable to select the best saving rate (or, indeed, to determine whether a constant saving rate is desirable) until we specify a detailed objective function, as we do in the next chapter. We can, however, argue in the present context that a saving rate that exceeds s_{gold} forever is inefficient because higher quantities of per capita consumption could be obtained at all points in time by reducing the saving rate.

Consider an economy, such as the one described by the saving rate s_2 in figure 1.3, for which $s_2 > s_{\text{gold}}$, so that $k_2^* > k_{\text{gold}}^*$ and $c_2^* < c_{\text{gold}}$. Imagine that, starting from the steady state, the saving rate is reduced permanently to s_{gold}. Figure 1.3 shows that per capita consumption, c—given by the vertical distance between the $f(k)$ and $s_{\text{gold}} \cdot f(k)$ curves—initially increases by a discrete amount. Then the level of c falls monotonically during the transition[14] toward its new steady-state value, c_{gold}. Since $c_2^* < c_{\text{gold}}$, we conclude that c exceeds its previous value, c_2^*, at all transitional dates, as well as in the new steady state. Hence, when $s > s_{\text{gold}}$, the economy is oversaving in the sense that per capita consumption at all points in time could be raised by lowering the saving rate. An economy that oversaves is said to be *dynamically inefficient,* because the path of per capita consumption lies below feasible alternative paths at all points in time.

If $s < s_{\text{gold}}$—as in the case of the saving rate s_1 in figure 1.3—then the steady-state amount of per capita consumption can be increased by raising the saving rate. This rise in the saving rate would, however, reduce c currently and during part of the transition period. The outcome will therefore be viewed as good or bad depending on how households weigh today's consumption against the path of future consumption. We cannot judge the desirability of an increase in the saving rate in this situation until we make specific assumptions about how agents discount the future. We proceed along these lines in the next chapter.

1.2.6 Transitional Dynamics

The long-run growth rates in the Solow–Swan model are determined entirely by exogenous elements—in the steady state, the per capita quantities k, y, and c do not grow and the aggregate variables K, Y, and C grow at the exogenous rate of population growth n. Hence, the main substantive conclusions about the long run are that steady-state growth rates are independent of the saving rate or the level of technology. The model does, however, have more interesting implications about transitional dynamics. This transition shows how an economy's per capita income converges toward its own steady-state value and to the per capita incomes of other economies.

Division of both sides of equation (1.13) by k implies that the growth rate of k is given by

$$\gamma_k \equiv \dot{k}/k = s \cdot f(k)/k - (n + \delta) \tag{1.23}$$

where we have used the notation γ_z to represent the growth rate of variable z, notation that we will use throughout the book. Note that, at all points in time, the growth rate of the level of a variable equals the per capita growth rate plus the exogenous rate of population growth n, for example,

$$\dot{K}/K = \dot{k}/k + n$$

For subsequent purposes, we shall find it convenient to focus on the growth rate of k, as given in equation (1.23).

14. In the next subsection we analyze the transitional dynamics of the model.

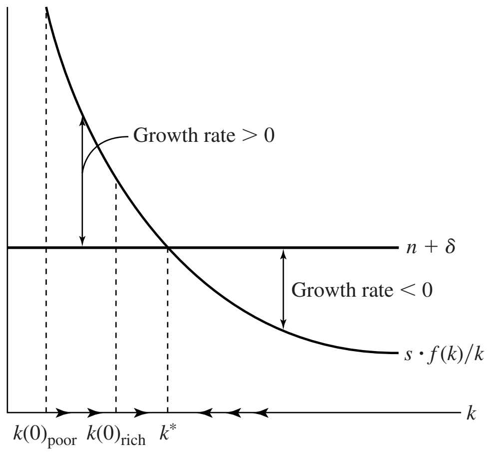

Figure 1.4
Dynamics of the Solow–Swan model. The growth rate of k is given by the vertical distance between the saving curve, $s \cdot f(k)/k$, and the effective depreciation line, $n + \delta$. If $k < k^*$, the growth rate of k is positive, and k increases toward k^*. If $k > k^*$, the growth rate is negative, and k falls toward k^*. Thus, the steady-state capital per person, k^*, is stable. Note that, along a transition from an initially low capital per person, the growth rate of k declines monotonically toward zero. The arrows on the horizontal axis indicate the direction of movement of k over time.

Equation (1.23) says that $\dot{k}/k$ equals the difference between two terms. The first term, $s \cdot f(k)/k$, we call the *saving curve* and the second term, $(n + \delta)$, the *depreciation curve*. We plot the two curves versus k in figure 1.4. The saving curve is downward sloping;[15] it asymptotes to infinity at $k = 0$ and approaches 0 as k tends to infinity.[16] The depreciation curve is a horizontal line at $n + \delta$. The vertical distance between the saving curve and the depreciation line equals the growth rate of capital per person (from equation [1.23]), and the crossing point corresponds to the steady state. Since $n + \delta > 0$ and $s \cdot f(k)/k$ falls monotonically from infinity to 0, the saving curve and the depreciation line intersect once and only once. Hence, the steady-state capital-labor ratio $k^* > 0$ exists and is unique.

Figure 1.4 shows that, to the left of the steady state, the $s \cdot f(k)/k$ curve lies above $n + \delta$. Hence, the growth rate of k is positive, and k rises over time. As k increases, $\dot{k}/k$ declines and approaches 0 as k approaches k^*. (The saving curve gets closer to the depreciation

15. The derivative of $f(k)/k$ with respect to k equals $-[f(k)/k - f'(k)]/k$. The expression in brackets equals the marginal product of labor, which is positive. Hence, the derivative is negative.

16. Note that $\lim_{k\to 0}[s \cdot f(k)/k] = 0/0$. We can apply l'Hôpital's rule to get $\lim_{k\to 0}[s \cdot f(k)/k] = \lim_{k\to 0}[s \cdot f'(k)] = \infty$, from the Inada condition. Similarly, the Inada condition $\lim_{k\to\infty}[f'(k)] = 0$ implies $\lim_{k\to\infty}[s \cdot f(k)/k] = 0$.

line as k gets closer to k^*; hence, $\dot{k}/k$ falls.) The economy tends asymptotically toward the steady state in which k—and, hence, y and c—do not change.

The reason behind the declining growth rates along the transition is the existence of diminishing returns to capital: when k is relatively low, the average product of capital, $f(k)/k$, is relatively high. By assumption, households save and invest a constant fraction, s, of this product. Hence, when k is relatively low, the gross investment per unit of capital, $s \cdot f(k)/k$, is relatively high. Capital per worker, k, effectively depreciates at the constant rate $n + \delta$. Consequently, the growth rate, $\dot{k}/k$, is also relatively high.

An analogous argument demonstrates that if the economy starts above the steady state, $k(0) > k^*$, then the growth rate of k is negative, and k falls over time. (Note from figure 1.4 that, for $k > k^*$, the $n + \delta$ line lies above the $s \cdot f(k)/k$ curve, and, hence, $\dot{k}/k < 0$.) The growth rate increases and approaches 0 as k approaches k^*. Thus, the system is globally stable: for any initial value, $k(0) > 0$, the economy converges to its unique steady state, $k^* > 0$.

We can also study the behavior of output along the transition. The growth rate of output per capita is given by

$$\dot{y}/y = f'(k) \cdot \dot{k}/f(k) = [k \cdot f'(k)/f(k)] \cdot (\dot{k}/k) \tag{1.24}$$

The expression in brackets on the far right is the *capital share,* that is, the share of the rental income on capital in total income.[17]

Equation (1.24) shows that the relation between $\dot{y}/y$ and $\dot{k}/k$ depends on the behavior of the capital share. In the Cobb–Douglas case (equation [1.11]), the capital share is the constant α, and $\dot{y}/y$ is the fraction α of $\dot{k}/k$. Hence, the behavior of $\dot{y}/y$ mimics that of $\dot{k}/k$.

More generally, we can substitute for $\dot{k}/k$ from equation (1.23) into equation (1.24) to get

$$\dot{y}/y = s \cdot f'(k) - (n + \delta) \cdot \mathrm{Sh}(k) \tag{1.25}$$

where $\mathrm{Sh}(k) \equiv k \cdot f'(k)/f(k)$ is the capital share. If we differentiate with respect to k and combine terms, we get

$$\partial(\dot{y}/y)/\partial k = \left[\frac{f''(k) \cdot k}{f(k)}\right] \cdot (\dot{k}/k) - \frac{(n + \delta) f'(k)}{f(k)} \cdot [1 - \mathrm{Sh}(k)]$$

Since $0 < \mathrm{Sh}(k) < 1$, the last term on the right-hand side is negative. If $\dot{k}/k \geq 0$, the first term

17. We showed before that, in a competitive market equilibrium, each unit of capital receives a rental equal to its marginal product, $f'(k)$. Hence, $k \cdot f'(k)$ is the income per person earned by owners of capital, and $k \cdot f'(k)/f(k)$—the term in brackets—is the share of this income in total income per person.

on the right-hand side is nonpositive, and, hence, $\partial(\dot{y}/y)/\partial k < 0$. Thus, $\dot{y}/y$ necessarily falls as k rises (and therefore as y rises) in the region in which $\dot{k}/k \geq 0$, that is, if $k \leq k^*$. If $\dot{k}/k < 0$ ($k > k^*$), the sign of $\partial(\dot{y}/y)/\partial k$ is ambiguous for a general form of the production function, $f(k)$. However, if the economy is close to its steady state, the magnitude of $\dot{k}/k$ will be small, and $\partial(\dot{y}/y)/\partial k < 0$ will surely hold even if $k > k^*$.

In the Solow–Swan model, which assumes a constant saving rate, the level of consumption per person is given by $c = (1 - s) \cdot y$. Hence, the growth rates of consumption and income per capita are identical at all points in time, $\dot{c}/c = \dot{y}/y$. Consumption, therefore, exhibits the same dynamics as output.

1.2.7 Behavior of Input Prices During the Transition

We showed before that the Solow–Swan framework is consistent with a competitive market economy in which firms maximize profits and households choose to save a constant fraction of gross income. It is interesting to study the behavior of wages and interest rates along the transition as the capital stock increases toward the steady state. We showed that the interest rate equals the marginal product of capital minus the constant depreciation rate, $r = f'(k) - \delta$. Since the interest rate depends on the marginal product of capital, which depends on the capital stock per person, the interest rate moves during the transition as capital changes. The neoclassical production function exhibits diminishing returns to capital, $f''(k) < 0$, so the marginal product of capital declines as capital grows. It follows that the interest rate declines monotonically toward its steady-state value, given by $r^* = f'(k^*) - \delta$.

We also showed that the competitive wage rate was given by $w = f(k) - k \cdot f'(k)$. Again, the wage rate moves as capital increases. To see the behavior of the wage rate, we can take the derivative of w with respect to k to get

$$\frac{\partial w}{\partial k} = f'(k) - f'(k) - k \cdot f''(k) = -k \cdot f''(k) > 0$$

The wage rate, therefore, increases monotonically as the capital stock grows. In the steady state, the wage rate is given by $w^* = f(k^*) - k^* \cdot f'(k^*)$.

The behavior of wages and interest rates can be seen graphically in figure 1.5. The curve shown in the figure is again the production function, $f(k)$. The income per worker received by individual households is given by

$$y = w + Rk \tag{1.26}$$

where $R = r + \delta$ is the rental price of capital. Once the interest rate and the wage rate are determined, y is a linear function of k, with intercept w and slope R.

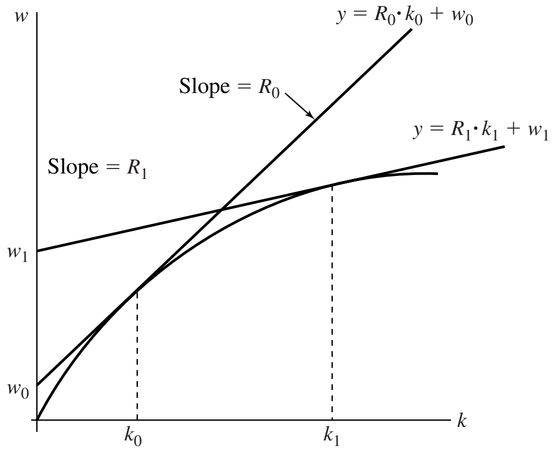

Figure 1.5
Input prices during the transition. At k_0, the straight line that is tangent to the production function has a slope that equals the rental price R_0 and an intercept that equals the wage rate w_0. As k rises toward k_1, the rental price falls toward R_1, and the wage rate rises toward w_1.

Of course, R depends on k through the marginal productivity condition, $f'(k) = R = r + \delta$. Therefore, R, the slope of the income function in equation (1.26), must equal the slope of $f(k)$ at the specified value of k. The figure shows two values, k_0 and k_1. The income functions at these two values are given by straight lines that are tangent to $f(k)$ at k_0 and k_1, respectively. As k rises during the transition, the figure shows that the slope of the tangent straight line declines from R_0 to R_1. The figure also shows that the intercept—which equals w—rises from w_0 to w_1.

1.2.8 Policy Experiments

Suppose that the economy is initially in a steady-state position with the capital per person equal to k_1^*. Imagine that the saving rate rises permanently from s_1 to a higher value s_2, possibly because households change their behavior or the government introduces some policy that raises the saving rate. Figure 1.6 shows that the $s \cdot f(k)/k$ schedule shifts to the right. Hence, the intersection with the $n + \delta$ line also shifts to the right, and the new steady-state capital stock, k_2^*, exceeds k_1^*.

How does the economy adjust from k_1^* to k_2^*? At $k = k_1^*$, the gap between the $s_1 \cdot f(k)/k$ curve and the $n + \delta$ line is positive; that is, saving is more than enough to generate an increase in k. As k increases, its growth rate falls and approaches 0 as k approaches k_2^*. The result, therefore, is that a permanent increase in the saving rate generates temporarily

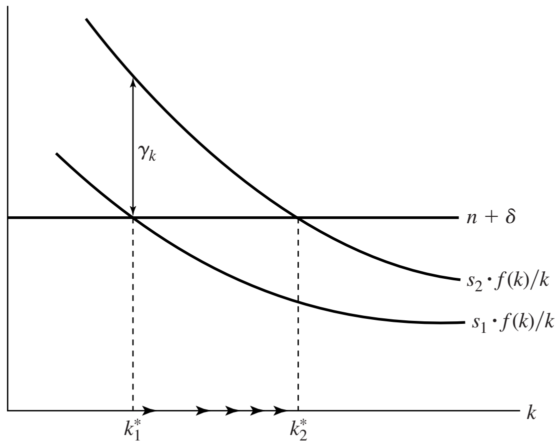

Figure 1.6
Effects from an increase in the saving rate. Starting from the steady-state capital per person k_1^*, an increase in s from s_1 to s_2 shifts the $s \cdot f(k)/k$ curve to the right. At the old steady state, investment exceeds effective depreciation, and the growth rate of k becomes positive. Capital per person rises until the economy approaches its new steady state at $k_2^* > k_1^*$.

positive per capita growth rates. In the long run, the levels of k and y are permanently higher, but the per capita growth rates return to zero.

The positive transitional growth rates may suggest that the economy could grow forever by raising the saving rate over and over again. One problem with this line of reasoning is that the saving rate is a fraction, a number between zero and one. Since people cannot save more than everything, the saving rate is bounded by one. Notice that, even if people could save all their income, the saving curve would still cross the depreciation line and, as a result, long-run per capita growth would stop.[18] The reason is that the workings of diminishing returns to capital eventually bring the economy back to the zero-growth steady state. Therefore, we can now answer the question that motivated the beginning of this chapter: "Can income per capita grow forever by simply saving and investing physical capital?" If the production function is neoclassical, the answer is "no."

We can also assess permanent changes in the growth rate of population, n. These changes could reflect shifts of household behavior or changes in government policies that influence fertility. A decrease in n shifts the depreciation line downward, so that the steady-state level of capital per worker would be larger. However, the long-run growth rate of capital per person would remain at zero.

18. Before reaching $s = 1$, the economy would reach s_{gold}, so that further increases in saving rates would put the economy in the dynamically inefficient region.

A permanent, once-and-for-all improvement in the level of the technology has similar, temporary effects on the per capita growth rates. If the production function $f(k)$ shifts upward in a proportional manner, then the saving curve shifts upward, just as in figure 1.6. Hence, $\dot{k}/k$ again becomes positive temporarily. In the long run, the permanent improvement in technology generates higher levels of k and y but no changes in the per capita growth rates. The key difference between improvements in knowledge and increases in the saving rate is that improvements in knowledge are not bounded. That is, the production function can shift over and over again because, in principle, there are no limits to human knowledge. The saving rate, however, is physically bounded by one. It follows that, if we want to generate growth in long-run per capita income and consumption within the neoclassical framework, growth must come from technological progress rather than from physical capital accumulation.

We observed before (note 3) that differences in government policies and institutions can amount to variations in the level of the technology. For example, high tax rates on capital income, failures to protect property rights, and distorting government regulations can be economically equivalent to a poorer level of technology. However, it is probably infeasible to achieve perpetual growth through an unending sequence of improvements in government policies and institutions. Therefore, in the long run, sustained growth would still depend on technological progress.

1.2.9 An Example: Cobb–Douglas Technology

We can illustrate the results for the case of a Cobb–Douglas production function (equation [1.11]). The steady-state capital-labor ratio is determined from equation (1.20) as

$$k^* = [sA/(n+\delta)]^{1/(1-\alpha)} \tag{1.27}$$

Note that, as we saw graphically for a more general production function $f(k)$, k^* rises with the saving rate s and the level of technology A, and falls with the rate of population growth n and the depreciation rate δ. The steady-state level of output per capita is given by

$$y^* = A^{1/(1-\alpha)} \cdot [s/(n+\delta)]^{\alpha/(1-\alpha)}$$

Thus y^* is a positive function of s and A, and a negative function of n and δ.

Along the transition, the growth rate of k is given from equation (1.23) by

$$\dot{k}/k = sAk^{-(1-\alpha)} - (n+\delta) \tag{1.28}$$

If $k(0) < k^*$, then $\dot{k}/k$ in equation (1.28) is positive. This growth rate declines as k rises and approaches 0 as k approaches k^*. Since equation (1.24) implies $\dot{y}/y = \alpha \cdot (\dot{k}/k)$, the behavior of $\dot{y}/y$ mimics that of $\dot{k}/k$. In particular, the lower $y(0)$, the higher $\dot{y}/y$.

A Closed-Form Solution It is interesting to notice that, when the production function is Cobb–Douglas and the saving rate is constant, it is possible to get a closed-form solution for the exact time path of k. Equation (1.28) can be written as

$$\dot{k} \cdot k^{-\alpha} + (n + \delta) \cdot k^{1-\alpha} = sA$$

If we define $v \equiv k^{1-\alpha}$, we can transform the equation to

$$\left(\frac{1}{1-\alpha}\right) \cdot \dot{v} + (n + \delta) \cdot v = sA$$

which is a first-order, linear differential equation in v. The solution to this equation is

$$v \equiv k^{1-\alpha} = \frac{sA}{(n+\delta)} + \left\{[k(0)]^{1-\alpha} - \frac{sA}{(n+\delta)}\right\} \cdot e^{-(1-\alpha)\cdot(n+\delta)\cdot t}$$

The last term is an exponential function with exponent equal to $-(1-\alpha) \cdot (n + \delta)$. Hence, the gap between $k^{1-\alpha}$ and its steady-state value, $sA/(n+\delta)$, vanishes exactly at the constant rate $(1 - \alpha) \cdot (n + \delta)$.

1.2.10 Absolute and Conditional Convergence

The fundamental equation of the Solow–Swan model (equation [1.23]) implies that the derivative of $\dot{k}/k$ with respect to k is negative:

$$\partial(\dot{k}/k)/\partial k = s \cdot [f'(k) - f(k)/k]/k < 0$$

Other things equal, smaller values of k are associated with larger values of $\dot{k}/k$. An important question arises: does this result mean that economies with lower capital per person tend to grow faster in per capita terms? In other words, does there tend to be *convergence* across economies?

To answer these questions, consider a group of closed economies (say, isolated regions or countries) that are structurally similar in the sense that they have the same values of the parameters s, n, and δ and also have the same production function $f(\cdot)$. Thus, the economies have the same steady-state values k^* and y^*. Imagine that the only difference among the economies is the initial quantity of capital per person $k(0)$. These differences in starting values could reflect past disturbances, such as wars or transitory shocks to production functions. The model then implies that the less-advanced economies—with lower values of $k(0)$ and $y(0)$—have higher growth rates of k and, in the typical case, also higher growth rates of y.[19]

19. This conclusion is unambiguous if the production function is Cobb–Douglas, if $k \leq k^*$, or if k is only a small amount above k^*.

Figure 1.4 distinguished two economies, one with the low initial value, $k(0)_{\text{poor}}$, and the other with the high initial value, $k(0)_{\text{rich}}$. Since each economy has the same underlying parameters, the dynamics of k are determined in each case by the same $s \cdot f(k)/k$ and $n+\delta$ curves. Hence, the growth rate $\dot{k}/k$ is unambiguously higher for the economy with the lower initial value, $k(0)_{\text{poor}}$. This result implies a form of convergence: regions or countries with lower starting values of the capital-labor ratio have higher per capita growth rates $\dot{k}/k$, and tend thereby to catch up or converge to those with higher capital-labor ratios.

The hypothesis that poor economies tend to grow faster per capita than rich ones—without conditioning on any other characteristics of economies—is referred to as *absolute convergence*. This hypothesis receives only mixed reviews when confronted with data on groups of economies. We can look, for example, at the growth experience of a broad cross section of countries over the period 1960 to 2000. Figure 1.7 plots the average annual growth rate of real per capita GDP against the log of real per capita GDP at the start of the period, 1960, for 114 countries. The growth rates are actually positively correlated with the initial position; that is, there is some tendency for the initially richer countries to grow faster in per capita terms. Thus, this sample rejects the hypothesis of absolute convergence.

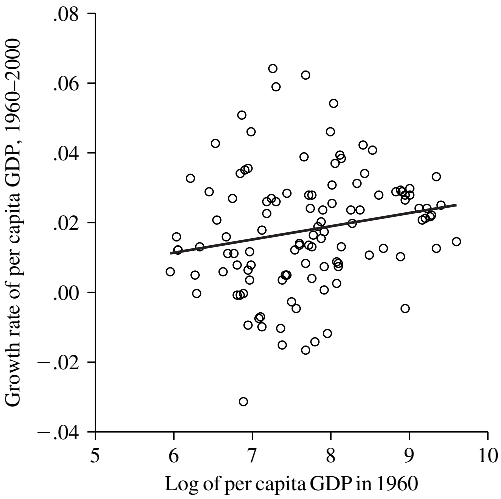

Figure 1.7
Convergence of GDP across countries: Growth rate versus initial level of real per capita GDP for 114 countries. For a sample of 114 countries, the average growth rate of GDP per capita from 1960 to 2000 (shown on the vertical axis) has little relation with the 1960 level of real per capita GDP (shown on the horizontal axis). The relation is actually slightly positive. Hence, absolute convergence does not apply for a broad cross section of countries.

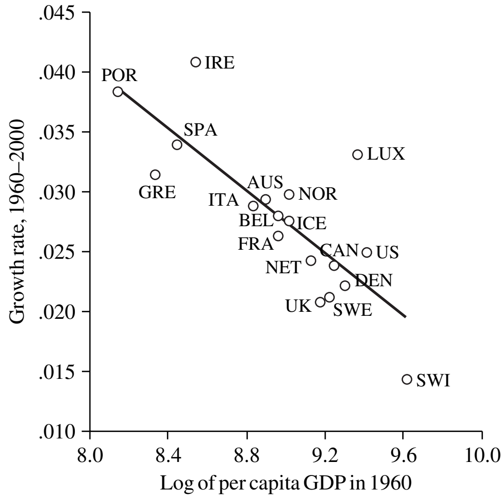

Figure 1.8
Convergence of GDP across OECD countries: Growth rate versus initial level of real per capita GDP for 18 OECD countries. If the sample is limited to 18 original OECD countries (from 1961), the average growth rate of real per capita GDP from 1960 to 2000 is negatively related to the 1960 level of real per capita GDP. Hence, absolute convergence applies for these OECD countries.

The hypothesis fares better if we examine a more homogeneous group of economies. Figure 1.8 shows the results if we limit consideration to 18 relatively advanced countries that were members of the Organization for Economic Cooperation and Development (OECD) from the start of the organization in 1961.[20] In this case, the initially poorer countries did experience significantly higher per capita growth rates.

This type of result becomes more evident if we consider an even more homogeneous group, the continental U.S. states, each viewed as a separate economy. Figure 1.9 plots the growth rate of per capita personal income for each state from 1880 to 2000 against the log of per capita personal income in 1880.[21] Absolute convergence—the initially poorer states growing faster in per capita terms—holds clearly in this diagram.

We can accommodate the theory to the empirical observations on convergence if we allow for heterogeneity across economies, in particular, if we drop the assumption that all economies have the same parameters, and therefore, the same steady-state positions. If the

20. Germany is omitted because of missing data, and Turkey is omitted because it was not an advanced economy in 1960.

21. There are 47 observations on U.S. states or territories. Oklahoma is omitted because 1880 preceded the Oklahoma land rush, and the data are consequently unavailable.

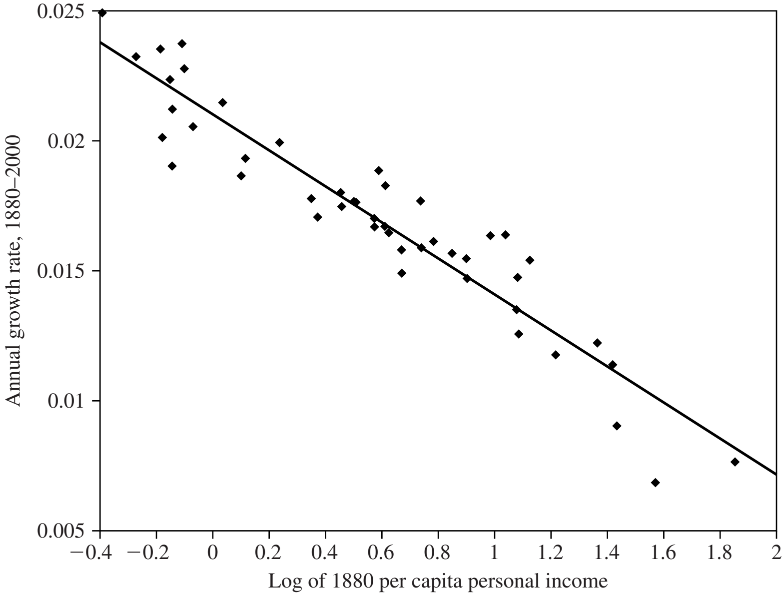

Figure 1.9
Convergence of personal income across U.S. states: 1880 personal income and income growth from 1880 to 2000. The relation between the growth rate of per capita personal income from 1880 to 2000 (shown on the vertical axis) is negatively related to the level of per capita income in 1880 (shown on the horizontal axis). Thus absolute convergence holds for the states of the United States.

steady states differ, we have to modify the analysis to consider a concept of *conditional convergence*. The main idea is that an economy grows faster the further it is from its own steady-state value.

We illustrate the concept of conditional convergence in figure 1.10 by considering two economies that differ in only two respects: first, they have different initial stocks of capital per person, $k(0)_{\text{poor}} < k(0)_{\text{rich}}$, and second, they have different saving rates, $s_{\text{poor}} \neq s_{\text{rich}}$. Our previous analysis implies that differences in saving rates generate differences in the same direction in the steady-state values of capital per person, that is, $k^*_{\text{poor}} \neq k^*_{\text{rich}}$. [In figure 1.10, these steady-state values are determined by the intersection of the $s_i \cdot f(k)/k$ curves with the common $n + \delta$ line.] We consider the case in which $s_{\text{poor}} < s_{\text{rich}}$ and, hence, $k^*_{\text{poor}} < k^*_{\text{rich}}$ because these differences likely explain why $k(0)_{\text{poor}} < k(0)_{\text{rich}}$ applies at the initial date. (It is also true empirically, as discussed in the introduction, that countries with higher levels of real per capita GDP tend to have higher saving rates.)

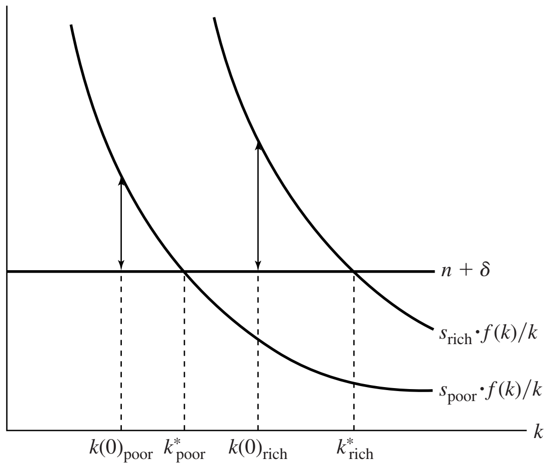

Figure 1.10
Conditional convergence. If a rich economy has a higher saving rate than a poor economy, the rich economy may be proportionately further from its steady-state position. In this case, the rich economy would be predicted to grow faster per capita than the poor economy; that is, absolute convergence would not hold.

The question is, Does the model predict that the poor economy will grow faster than the rich one? If they have the same saving rate, then the per capita growth rate—the distance between the $s \cdot f(k)/k$ curve and the $n+\delta$ line—would be higher for the poor economy, and $(\dot{k}/k)_{\text{poor}} > (\dot{k}/k)_{\text{rich}}$ would apply. However, if the rich economy has a higher saving rate, as in figure 1.10, then $(\dot{k}/k)_{\text{poor}} < (\dot{k}/k)_{\text{rich}}$ might hold, so that the rich economy grows faster. The intuition is that the low saving rate of the poor economy offsets its higher average product of capital as a determinant of economic growth. Hence, the poor economy may grow at a slower rate than the rich one.

The neoclassical model does predict that each economy converges to its own steady state and that the speed of this convergence relates inversely to the distance from the steady state. In other words, the model predicts conditional convergence in the sense that a lower starting value of real per capita income tends to generate a higher per capita growth rate, once we control for the determinants of the steady state.

Recall that the steady-state value, k^*, depends on the saving rate, s, and the level of the production function, $f(\cdot)$. We have also mentioned that government policies and institutions can be viewed as additional elements that effectively shift the position of the production function. The findings on conditional convergence suggest that we should hold constant these determinants of k^* to isolate the predicted inverse relationship between growth rates and initial positions.

Algebraically, we can illustrate the concept of conditional convergence by returning to the formula for $\dot{k}/k$ in equation (1.23). One of the determinants of $\dot{k}/k$ is the saving rate s. We can use the steady-state condition from equation (1.20) to express s as follows:

$$s = (n + \delta) \cdot k^*/f(k^*)$$

If we replace s by this expression in equation (1.23), then $\dot{k}/k$ can be expressed as

$$\dot{k}/k = (n + \delta) \cdot \left[\frac{f(k)/k}{f(k^*)/k^*} - 1 \right] \tag{1.29}$$

Equation (1.29) is consistent with $\dot{k}/k = 0$ when $k = k^*$. For given k^*, the formula implies that a reduction in k, which raises the average product of capital, $f(k)/k$, increases $\dot{k}/k$. But a lower k matches up with a higher $\dot{k}/k$ only if the reduction is relative to the steady-state value, k^*. In particular, $f(k)/k$ must be high relative to the steady-state value, $f(k^*)/k^*$. Thus a poor country would not be expected to grow rapidly if its steady-state value, k^*, is as low as its current value, k.

In the case of a Cobb–Douglas technology, the saving rate can be written as

$$s = \frac{(n + \delta)}{A} \cdot k^{*(1-\alpha)}$$

which we can substitute into equation (1.23) to get

$$\dot{k}/k = (n + \delta) \cdot \left[\left(\frac{k}{k^*} \right)^{\alpha - 1} - 1 \right] \tag{1.30}$$

We see that the growth rate of capital, k, depends on the ratio k/k^*; that is, it depends on the distance between the current and steady-state capital-labor ratio.

The result in equation (1.29) suggests that we should look empirically at the relation between the per capita growth rate, $\dot{y}/y$, and the starting position, $y(0)$, after holding fixed variables that account for differences in the steady-state position, y^*. For a relatively homogeneous group of economies, such as the U.S. states, the differences in steady-state positions may be minor, and we would still observe the convergence pattern shown in figure 1.9. For a broad cross section of 114 countries, however, as shown in figure 1.7, the differences in steady-state positions are likely to be substantial. Moreover, the countries with low starting levels, $y(0)$, are likely to be in this position precisely because they have low steady-state values, y^*, perhaps because of chronically low saving rates or persistently bad government policies that effectively lower the level of the production function. In other words, the per capita growth rate may have little correlation with $\log[y(0)]$, as in figure 1.7, because $\log[y(0)]$ is itself uncorrelated with the gap from the steady state, $\log[y(0)/y^*]$. The perspective of conditional convergence indicates that this gap is the variable that matters for the subsequent per capita growth rate.

We show in chapter 12 that the inclusion of variables that proxy for differences in steady-state positions makes a major difference in the results for the broad cross section of countries. When these additional variables are held constant, the relation between the per capita growth rate and the log of initial real per capita GDP becomes significantly negative, as predicted by the neoclassical model. In other words, the cross-country data support the hypothesis of conditional convergence.

1.2.11 Convergence and the Dispersion of Per Capita Income

The concept of convergence considered thus far is that economies with lower levels of per capita income (expressed relative to their steady-state levels of per capita income) tend to grow faster in per capita terms. This behavior is often confused with an alternative meaning of convergence, that the dispersion of real per capita income across a group of economies or individuals tends to fall over time.[22] We show now that, even if absolute convergence holds in our sense, the dispersion of per capita income need not decline over time.

Suppose that absolute convergence holds for a group of economies $i = 1, \ldots, N$, where N is a large number. In discrete time, corresponding for example to annual data, the real per capita income for economy i can then be approximated by the process

$$\log(y_{it}) = a + (1 - b) \cdot \log(y_{i,t-1}) + u_{it} \tag{1.31}$$

where a and b are constants, with $0 < b < 1$, and u_{it} is a disturbance term. The condition $b > 0$ implies absolute convergence because the annual growth rate, $\log(y_{it}/y_{i,t-1})$, is inversely related to $\log(y_{i,t-1})$. A higher coefficient b corresponds to a greater tendency toward convergence.[23] The disturbance term picks up temporary shocks to the production function, the saving rate, and so on. We assume that u_{it} has zero mean, the same variance σ_u^2 for all economies, and is independent over time and across economies.

One measure of the dispersion or inequality of per capita income is the sample variance of the $\log(y_{it})$:

$$D_t \equiv \frac{1}{N} \cdot \sum_{i=1}^{N} [\log(y_{it}) - \mu_t]^2$$

22. See Sala-i-Martin (1990) and Barro and Sala-i-Martin (1992a) for further discussion of the two concepts of convergence.

23. The condition $b < 1$ rules out a leapfrogging or overshooting effect, whereby an economy that starts out behind another economy would be predicted systematically to get ahead of the other economy at some future date. This leapfrogging effect cannot occur in the neoclassical model but can arise in some models of technological adaptation that we discuss in chapter 8.

where μ_t is the sample mean of the $\log(y_{it})$. If there are a large number N of observations, the sample variance is close to the population variance, and we can use equation (1.31) to derive the evolution of D_t over time:

$$D_t \approx (1-b)^2 \cdot D_{t-1} + \sigma_u^2$$

This first-order difference equation for dispersion has a steady state given by

$$D^* = \sigma_u^2/[1-(1-b)^2]$$

Hence, the steady-state dispersion falls with b (the strength of the convergence effect) but rises with the variance σ_u^2 of the disturbance term. In particular, $D^* > 0$ even if $b > 0$, as long as $\sigma_u^2 > 0$.

The evolution of D_t can be expressed as

$$D_t = D^* + (1-b)^2 \cdot (D_{t-1} - D^*) = D^* + (1-b)^{2t} \cdot (D_0 - D^*) \tag{1.32}$$

where D_0 is the dispersion at time 0. Since $0 < b < 1$, D_t monotonically approaches its steady-state value, D^*, over time. Equation (1.32) implies that D_t rises or falls over time depending on whether D_0 begins below or above the steady-state value.[24] Note especially that a rising dispersion is consistent with absolute convergence ($b > 0$).

These results about convergence and dispersion are analogous to Galton's fallacy about the distribution of heights in a population (see Quah, 1993, and Hart, 1995, for discussions). The observation that heights in a family tend to regress toward the mean across generations (a property analogous to our convergence concept for per capita income) does not imply that the dispersion of heights across the full population (a measure that parallels the dispersion of per capita income across economies) tends to narrow over time.

1.2.12 Technological Progress

Classification of Inventions We have assumed thus far that the level of technology is constant over time. As a result, we found that all per capita variables were constant in the long run. This feature of the model is clearly unrealistic; in the United States, for example, the average per capita growth rate has been positive for over two centuries. In the absence of technological progress, diminishing returns would have made it impossible to maintain per capita growth for so long just by accumulating more capital per worker. The neoclassical economists of the 1950s and 1960s recognized this problem and amended the basic model

24. We could extend the model by allowing for temporary shocks to σ_u^2 or for major disturbances like wars or oil shocks that affect large subgroups of economies in a common way. In this extended model, the dispersion could depart from the deterministic path that we derived; for example, D_t could rise in some periods even if D_0 began above its steady-state value.

to allow the technology to improve over time. These improvements provided an escape from diminishing returns and thus enabled the economy to grow in per capita terms in the long run. We now explore how the model works when we allow for such technological advances.

Although some discoveries are serendipitous, most technological improvements reflect purposeful activity, such as research and development (R&D) carried out in universities and corporate or government laboratories. This research is sometimes financed by private institutions and sometimes by governmental agencies, such as the National Science Foundation. Since the amount of resources devoted to R&D depends on economic conditions, the evolution of the technology also depends on these conditions. This relation will be the subject of our analysis in chapters 6–8. At present, we consider only the simpler case in which the technology improves exogenously.

The first issue is how to introduce exogenous technological progress into the model. This progress can take various forms. Inventions may allow producers to generate the same amount of output with either relatively less capital input or relatively less labor input, cases referred to as *capital-saving* or *labor-saving* technological progress, respectively. Inventions that do not save relatively more of either input are called *neutral* or *unbiased*.

The definition of neutral technological progress depends on the precise meaning of capital saving and labor saving. Three popular definitions are due to Hicks (1932), Harrod (1942), and Solow (1969).

Hicks says that a technological innovation is neutral (Hicks neutral) if the ratio of marginal products remains unchanged for a given capital-labor ratio. This property corresponds to a renumbering of the isoquants, so that Hicks-neutral production functions can be written as

$$Y = T(t) \cdot F(K, L) \tag{1.33}$$

where $T(t)$ is the index of the state of the technology, and $\dot{T}(t) \geq 0$.

Harrod defines an innovation as neutral (Harrod neutral) if the relative input shares, $(K \cdot F_K)/(L \cdot F_L)$, remain unchanged for a given capital-output ratio. Robinson (1938) and Uzawa (1961) showed that this definition implied that the production function took the form

$$Y = F[K, L \cdot T(t)] \tag{1.34}$$

where $T(t)$ is the index of the technology, and $\dot{T}(t) \geq 0$. This form is called *labor-augmenting* technological progress because it raises output in the same way as an increase in the stock of labor. (Notice that the technology factor, $T(t)$, appears in the production function as a multiple of L.)

Finally, Solow defines an innovation as neutral (Solow neutral) if the relative input shares, $(L \cdot F_L)/(K \cdot F_K)$, remain unchanged for a given labor/output ratio. This definition can be shown to imply a production function of the form

$$Y = F[K \cdot T(t), L] \tag{1.35}$$

where $T(t)$ is the index of the technology, and $\dot{T}(t) \geq 0$. Production functions of this form are called *capital augmenting* because a technological improvement increases production in the same way as an increase in the stock of capital.

The Necessity for Technological Progress to Be Labor Augmenting Suppose that we consider only constant rates of technological progress. Then, in the neoclassical growth model with a constant rate of population growth, only labor-augmenting technological change turns out to be consistent with the existence of a steady state, that is, with constant growth rates of the various quantities in the long run. This result is proved in the appendix to this chapter (section 1.5).

If we want to consider models that possess a steady state, we have to assume that technological progress takes the labor-augmenting form. Another approach, which would be substantially more complicated, would be to deal with models that lack steady states, that is, in which the various growth rates do not approach constants in the long run. However, one reason to stick with the simpler framework that possesses a steady state is that the long-term experiences of the United States and some other developed countries indicate that per capita growth rates can be positive and trendless over long periods of time (see chapter 12). This empirical phenomenon suggests that a useful theory would predict that per capita growth rates approach constants in the long run; that is, the model would possess a steady state.

If the production function is Cobb–Douglas, $Y = AK^{\alpha}L^{1-\alpha}$ in equation (1.11), then it is clear from inspection that the form of technological progress—augmenting A, K, or L—will not matter for the results (see the appendix for discussion). Thus, in the Cobb–Douglas case, we will be safe in assuming that technological progress is labor augmenting. Recall that the key property of the Cobb–Douglas function is that, in a competitive setting, the factor-income shares are constant. Thus, if factor-income shares are reasonably stable—as seems to be true for the U.S. economy but not for some others—we may be okay in regarding the production function as approximately Cobb–Douglas and, hence, in assuming that technogical progress is labor augmenting.

Another approach, when the production function is not Cobb–Douglas, is to derive the form of technological progress from a theory of technological change. Acemoglu (2002) takes this approach, using a variant of the model of endogenous technological change that we develop in chapter 6. He finds that, under some conditions, the form of technological progress would be asymptotically labor augmenting.

The Solow–Swan Model with Labor-Augmenting Technological Progress We assume now that the production function includes labor-augmenting technological progress, as shown in equation (1.34), and that the technology term, $T(t)$, grows at the constant rate x. The condition for the change in the capital stock is

$$\dot{K} = s \cdot F[K, L \cdot T(t)] - \delta K$$

If we divide both sides of this equation by L, we can derive an expression for the change in k over time:

$$\dot{k} = s \cdot F[k, T(t)] - (n + \delta) \cdot k \tag{1.36}$$

The only difference from equation (1.13) is that output per person now depends on the level of the technology, $T(t)$.

Divide both sides of equation (1.36) by k to compute the growth rate:

$$\dot{k}/k = s \cdot F[k, T(t)]/k - (n + \delta) \tag{1.37}$$

As in equation (1.23), $\dot{k}/k$ equals the difference between two terms, where the first term is the product of s and the average product of capital, and the second term is $n + \delta$. The only difference is that now, for given k, the average product of capital, $F[k, T(t)]/k$, increases over time because of the growth in $T(t)$ at the rate x. In terms of figure 1.4, the downward-sloping curve, $s \cdot F(\cdot)/k$, shifts continually to the right, and, hence, the level of k that corresponds to the intersection between this curve and the $n + \delta$ line also shifts continually to the right. We now compute the growth rate of k in the steady state.

By definition, the steady-state growth rate, $(\dot{k}/k)^*$, is constant. Since s, n, and δ are also constants, equation (1.37) implies that the average product of capital, $F[k, T(t)]/k$, is constant in the steady state. Because of constant returns to scale, the expression for the average product equals $F[1, T(t)/k]$ and is therefore constant only if k and $T(t)$ grow at the same rate, that is, $(\dot{k}/k)^* = x$.

Output per capita is given by

$$y = F[k, T(t)] = k \cdot F[1, T(t)/k]$$

Since k and $T(t)$ grow in the steady state at the rate x, the steady-state growth rate of y equals x. Moreover, since $c = (1 - s) \cdot y$, the steady-state growth rate of c also equals x.

To analyze the transitional dynamics of the model with technological progress, it will be convenient to rewrite the system in terms of variables that remain constant in the steady state. Since k and $T(t)$ grow in the steady state at the same rate, we can work with the ratio $\hat{k} \equiv k/T(t) = K/[L \cdot T(t)]$. The variable $L \cdot T(t) \equiv \hat{L}$ is often called the *effective amount of labor*—the physical quantity of labor, L, multiplied by its efficiency, $T(t)$. (The terminology *effective labor* is appropriate because the economy operates as if its labor input were $\hat{L}$.) The variable $\hat{k}$ is then the quantity of capital per unit of effective labor.

The quantity of output per unit of effective labor, $\hat{y} \equiv Y/[L \cdot T(t)]$, is given by

$$\hat{y} = F(\hat{k}, 1) \equiv f(\hat{k}) \tag{1.38}$$

Hence, we can again write the production function in intensive form if we replace y and k by $\hat{y}$ and $\hat{k}$, respectively. If we proceed as we did before to get equations (1.13) and (1.23), but now use the condition that $A(t)$ grows at the rate x, we can derive the dynamic equation for $\hat{k}$:

$$\dot{\hat{k}}/\hat{k} = s \cdot f(\hat{k})/\hat{k} - (x + n + \delta) \tag{1.39}$$

The only difference between equations (1.39) and (1.23), aside from the hats (ˆ), is that the last term on the right-hand side includes the parameter x. The term $x + n + \delta$ is now the effective depreciation rate for $\hat{k} \equiv K/\hat{L}$. If the saving rate, s, were zero, $\hat{k}$ would decline partly due to depreciation of K at the rate δ and partly due to growth of $\hat{L}$ at the rate $x + n$.

Following an argument similar to that of section 1.2.4, we can show that the steady-state growth rate of $\hat{k}$ is zero. The steady-state value $\hat{k}^*$ satisfies the condition

$$s \cdot f(\hat{k}^*) = (x + n + \delta) \cdot \hat{k}^* \tag{1.40}$$

The transitional dynamics of $\hat{k}$ are qualitatively similar to those of k in the previous model. In particular, we can construct a picture like figure 1.4 in which the horizontal axis involves $\hat{k}$, the downward-sloping curve is now $s \cdot f(\hat{k})/\hat{k}$, and the horizontal line is at the level $x + n + \delta$, rather than $n + \delta$. The new construction is shown in figure 1.11. We can use this figure, as we used figure 1.4 before, to assess the relation between the initial value, $\hat{k}(0)$, and the growth rate, $\dot{\hat{k}}/\hat{k}$.

In the steady state, the variables with hats—$\hat{k}$, $\hat{y}$, $\hat{c}$—are now constant. Therefore, the per capita variables—k, y, c—now grow in the steady state at the exogenous rate of technological progress, x.[25] The level variables—K, Y, C—grow accordingly in the steady state at the rate $n + x$, that is, the sum of population growth and technological change. Note that, as in the prior analysis that neglected technological progress, shifts to the saving rate or the level of the production function affect long-run levels—$\hat{k}^*$, $\hat{y}^*$, $\hat{c}^*$—but not steady-state growth rates. As before, these kinds of disturbances influence growth rates during the transition from an initial position, represented by $\hat{k}(0)$, to the steady-state value, $\hat{k}^*$.

25. We always have the condition $(1/\hat{k}) \cdot (d\hat{k}/dt) = \dot{k}/k - x$. Therefore, $(1/\hat{k}) \cdot (d\hat{k}/dt) = 0$ implies $\dot{k}/k = x$, and similarly for $\dot{y}/y$ and $\dot{c}/c$.

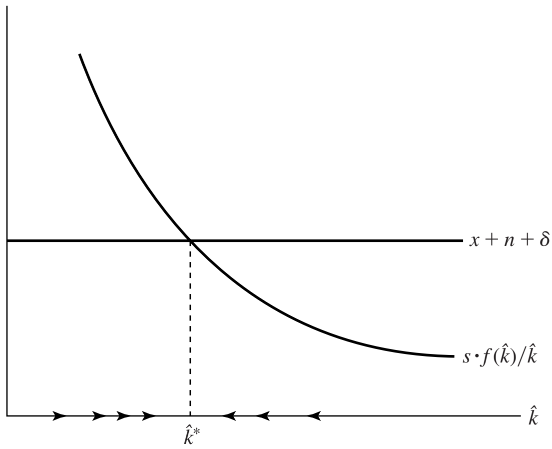

Figure 1.11
The Solow–Swan model with technological progress. The growth rate of capital per effective worker ($\hat{k} \equiv K/LT$) is given by the vertical distance between the $s \cdot f(\hat{k})/\hat{k}$ curve and the effective depreciation line, $x+n+\delta$. The economy is at a steady state when $\hat{k}$ is constant. Since T grows at the constant rate x, the steady-state growth rate of capital per person, k, also equals x.

1.2.13 A Quantitative Measure of the Speed of Convergence

It is important to know the speed of the transitional dynamics. If convergence is rapid, we can focus on steady-state behavior, because most economies would typically be close to their steady states. Conversely, if convergence is slow, economies would typically be far from their steady states, and, hence, their growth experiences would be dominated by the transitional dynamics.

We now provide a quantitative assessment of how fast the economy approaches its steady state for the case of a Cobb–Douglas production function, shown in equation (1.11). (We generalize later to a broader class of production functions.) We can use equation (1.39), with L replaced by $\hat{L}$, to determine the growth rate of $\hat{k}$ in the Cobb–Douglas case as

$$\dot{\hat{k}}/\hat{k} = sA \cdot (\hat{k})^{-(1-\alpha)} - (x+n+\delta) \tag{1.41}$$

The *speed of convergence,* β, is measured by how much the growth rate declines as the capital stock increases in a proportional sense, that is,

$$\beta \equiv -\frac{\partial(\dot{\hat{k}}/\hat{k})}{\partial \log \hat{k}} \tag{1.42}$$

Notice that we define β with a negative sign because the derivative is negative, so that β is positive.

To compute β, we have to rewrite the growth rate in equation (1.41) as a function of $\log(\hat{k})$:

$$\dot{\hat{k}}/\hat{k} = sA \cdot e^{-(1-\alpha)\cdot \log(\hat{k})} - (x + n + \delta) \tag{1.43}$$

We can take the derivative of equation (1.43) with respect to $\log(\hat{k})$ to get an expression for β:

$$\beta = (1-\alpha) \cdot sA \cdot (\hat{k})^{-(1-\alpha)} \tag{1.44}$$

Notice that the speed of convergence is not constant but, rather, declines monotonically as the capital stock increases toward its steady-state value. At the steady state, $sA \cdot (\hat{k})^{-(1-\alpha)} = (x + n + \delta)$ holds. Therefore, in the neighborhood of the steady state, the speed of convergence equals

$$\beta^* = (1-\alpha) \cdot (x + n + \delta) \tag{1.45}$$

During the transition to the steady state, the convergence rate, β, exceeds β^* but declines over time.

Another way to get the formula for β^* is to consider a log-linear approximation of equation (1.41) in the neighborhood of the steady state:

$$\dot{\hat{k}}/\hat{k} \cong -\beta^* \cdot [\log(\hat{k}/\hat{k}^*)] \tag{1.46}$$

where the coefficient β^* comes from a log-linearization of equation (1.41) around the steady state. The resulting coefficient can be shown to equal the right-hand side of equation (1.45). See the appendix at the end of this chapter (section 1.5) for the method of derivation of this log-linearization.

Before we consider further the implications of equation (1.45), we will show that it applies also to the growth rate of $\hat{y}$. For a Cobb–Douglas production function, shown in equation (1.11), we have

$$\dot{\hat{y}}/\hat{y} = \alpha \cdot (\dot{\hat{k}}/\hat{k})$$

$$\log(\hat{y}/\hat{y}^*) = \alpha \cdot \log(\hat{k}/\hat{k}^*)$$

If we substitute these formulas into equation (1.46), we get

$$\dot{\hat{y}}/\hat{y} \approx -\beta^* \cdot [\log(\hat{y}/\hat{y}^*)] \tag{1.47}$$

Hence, the convergence coefficient for $\hat{y}$ is the same as that for $\hat{k}$.

The term $\beta^* = (1-\alpha) \cdot (x + n + \delta)$ in equation (1.45) indicates how rapidly an economy's output per effective worker, $\hat{y}$, approaches its steady-state value, $\hat{y}^*$, in the neighborhood of the steady state. For example, if $\beta^* = 0.05$ per year, 5 percent of the gap between $\hat{y}$ and $\hat{y}^*$ vanishes in one year. The half-life of convergence—the time that it takes for half the initial gap to be eliminated—is thus about 14 years.[26] It would take about 28 years for three-quarters of the gap to vanish.

Consider what the theory implies quantitatively about the convergence coefficient, $\beta^* = (1 - \alpha) \cdot (x + n + \delta)$, in equation (1.45). One property is that the saving rate, s, does not affect β^*. This result reflects two offsetting forces that exactly cancel in the Cobb–Douglas case. First, given $\hat{k}$, a higher saving rate leads to greater investment and, therefore, to a faster speed of convergence. Second, a higher saving rate raises the steady-state capital intensity, $\hat{k}^*$, and thereby lowers the average product of capital in the vicinity of the steady state. This effect reduces the speed of convergence. The coefficient β^* is also independent of the overall level of efficiency of the economy, A. Differences in A, like differences in s, have two offsetting effects on the convergence speed, and these effects exactly cancel in the Cobb–Douglas case.

To see the quantitative implications of the parameters that enter into equation (1.45), consider the benchmark values $x = 0.02$ per year, $n = 0.01$ per year, and $\delta = 0.05$ per year. These values appear reasonable, for example, for the U.S. economy. The long-term growth rate of real GDP, which is about 2 percent per year, corresponds in the theory to the parameter x. The rate of population growth in recent decades is about 1 percent per year, and the measured depreciation rate for the overall stock of structures and equipment is around 5 percent per year.

For given values of the parameters x, n, and δ, the coefficient β^* in equation (1.45) is determined by the capital-share parameter, α. A conventional share for the gross income accruing to a narrow concept of physical capital (structures and equipment) is about $\frac{1}{3}$ (see Denison, 1962; Maddison, 1982; and Jorgenson, Gollop, and Fraumeni, 1987). If we use $\alpha = \frac{1}{3}$, equation (1.45) implies $\beta^* = 5.6$ percent per year, which implies a half-life of 12.5 years. In other words, if the capital share is $\frac{1}{3}$, the neoclassical model predicts relatively short transitions.

26. Equation (1.47) is a differential equation in $\log[\hat{y}(t)]$ with the solution

$$\log[\hat{y}(t)] = (1 - e^{-\beta^* t}) \cdot \log(\hat{y}^*) + e^{-\beta^* t} \cdot \log[\hat{y}(0)]$$

The time t for which $\log[\hat{y}(t)]$ is halfway between $\log[\hat{y}(0)]$ and $\log(\hat{y}^*)$ satisfies the condition $e^{-\beta^* t} = 1/2$. The half-life is therefore $\log(2)/\beta^* = 0.69/\beta^*$. Hence, if $\beta^* = 0.05$ per year, the half-life is 14 years.

In chapters 11 and 12 we argue that this predicted speed of convergence is much too high to accord with the empirical evidence. A convergence coefficient, β, in the range of 1.5 percent to 3.0 percent per year appears to fit better with the data. If $\beta^* = 2.0$ percent per year, the half-life is about 35 years, and the time needed to eliminate three-quarters of an initial gap from the steady-state position is about 70 years. In other words, convergence speeds that are consistent with the empirical evidence imply that the time required for substantial convergence is typically on the order of several generations.

To accord with an observed rate of convergence of about 2 percent per year, the neoclassical model requires a much higher capital-share coefficient. For example, the value $\alpha = 0.75$, together with the benchmark values for the other parameters, implies $\beta^* = 2.0$ percent per year. Although a capital share of 0.75 is too high for a narrow concept of physical capital, this share is reasonable for an expanded measure that also includes human capital.

An Extended Solow–Swan Model with Physical and Human Capital One way to increase the capital share is to add human capital to the model. Consider a Cobb–Douglas production function that uses physical capital, K, human capital, H,[27] and raw labor, L:

$$Y = AK^{\alpha}H^{\eta}[T(t)\cdot L]^{1-\alpha-\eta} \tag{1.48}$$

where $T(t)$ again grows at the exogenous rate x. Divide the production function by $T(t)\cdot L$ to get output per unit of effective labor:

$$\hat{y} = A\hat{k}^{\alpha}\hat{h}^{\eta} \tag{1.49}$$

Output can be used on a one-to-one basis for consumption or investment in either type of capital. Following Solow and Swan, we still assume that people consume a constant fraction, $1-s$, of their gross income, so the accumulation is given by

$$\dot{\hat{k}} + \dot{\hat{h}} = sA\hat{k}^{\alpha}\hat{h}^{\eta} - (\delta + n + x)\cdot(\hat{k} + \hat{h}) \tag{1.50}$$

where we have assumed that the two capital goods depreciate at the same constant rate.

The key question is how overall savings will be allocated between physical and human capital. It is reasonable to think that households will invest in the capital good that delivers the higher return, so that the two rates of return—and, hence, the two marginal products of capital—will have to be equated if both forms of investment are taking place. Therefore,

27. Chapters 4 and 5 discuss human capital in more detail.

we have the condition[28]

$$\alpha \cdot \frac{\hat{y}}{\hat{k}} - \delta = \eta \cdot \frac{\hat{y}}{\hat{h}} - \delta \tag{1.51}$$

The equality between marginal products implies a one-to-one relationship between physical and human capital:

$$\hat{h} = \frac{\eta}{\alpha} \cdot \hat{k} \tag{1.52}$$

We can use this relation to eliminate $\hat{h}$ from equation (1.50) to get

$$\dot{\hat{k}} = s\tilde{A}\hat{k}^{\alpha+\eta} - (\delta + n + x) \cdot \hat{k} \tag{1.53}$$

where $\tilde{A} \equiv (\frac{\eta^{\eta}\alpha^{(1-\eta)}}{\alpha+\eta}) \cdot A$ is a constant. Notice that this accumulation equation is the same as equation (1.41), except that the exponent on the capital stock per worker is now the sum of the physical and human capital shares, $\alpha + \eta$, instead of α. Using a derivation analogous to that of the previous section, we therefore get an expression for the convergence coefficient in the steady state:

$$\beta^* = (1 - \alpha - \eta) \cdot (\delta + n + x) \tag{1.54}$$

Jorgenson, Gollop, and Fraumeni (1987) estimate a human-capital share of between 0.4 and 0.5. With $\eta = 0.4$ and with the benchmark parameters of the previous section, including $\alpha = \frac{1}{3}$, the predicted speed of convergence would be $\beta^* = 0.021$. Thus, with a broad concept of capital that includes human capital, the Solow–Swan model can generate the rates of convergence that have been observed empirically.

Mankiw, Romer, and Weil (1992) use a production function analogous to equation (1.48). However, instead of making the Solow–Swan assumption that the overall gross saving rate is constant and exogenous, they assume that the investment rates in the two forms of capital are each constant and exogenous. For physical capital, the growth rate is therefore

$$\dot{\hat{k}} = s_k\tilde{A}\hat{k}^{\alpha-1}\hat{h}^{\eta} - (\delta + n + x) = s_k\tilde{A} \cdot e^{-(1-\alpha)\ln\hat{k}} \cdot e^{\eta\ln\hat{h}} - (\delta + n + x) \tag{1.55}$$

28. In a market setup, profit would be $\pi = AK_t^{\alpha}H_t^{\eta}(T_tL_t)^{1-\alpha-\eta} - R_kK - R_hH - wL$, where R_k and R_h are the rental rates of physical and human capital, respectively. The first-order conditions for the firm require that the marginal products of each of the capital goods be equalized to the rental rates, $R_k = \alpha\frac{\hat{y}}{\hat{k}}$ and $R_h = \eta\frac{\hat{y}}{\hat{h}}$. In an environment without uncertainty, like the one we are considering, physical capital, human capital, and loans are perfect substitutes as stores of value and, as a result, their net returns must be the same. In other words, $r = R_k - \delta = R_h - \delta$. Optimizing firms will, therefore, rent physical and human capital up to the point where their marginal products are equal.

where s_k is an exogenous constant. Similarly, for human capital, the growth rate is

$$\dot{\hat{h}} = s_h\tilde{A}\hat{k}^{\alpha}\hat{h}^{\eta-1} - (\delta+n+x) = s_h\tilde{A}\cdot e^{\alpha\ln\hat{k}}\cdot e^{-(1-\eta)\ln\hat{h}} - (\delta+n+x) \tag{1.56}$$

where s_h is another exogenous constant. A shortcoming of this approach is that the rates of return to physical and human capital are not equated.

The growth rate of $\hat{y}$ is a weighted average of the growth rates of the two inputs:

$$\dot{\hat{y}}/\hat{y} = \alpha\cdot(\dot{\hat{k}}/\hat{k}) + \eta\cdot(\dot{\hat{h}}/\hat{h})$$

If we use equations (1.55) and (1.56) and take a two-dimensional first-order Taylor-series expansion, we get

$$\begin{aligned}\dot{\hat{y}}/\hat{y} = {} & \big[\alpha s_k\tilde{A}\cdot e^{-(1-\alpha)\ln\hat{k}^*}\cdot e^{\eta\ln\hat{h}^*}\cdot[-(1-\alpha)] \\ & + \eta s_h\tilde{A}\cdot e^{\alpha\ln\hat{k}^*}\cdot e^{-(1-\eta)\ln\hat{h}^*}\cdot\alpha\big]\cdot(\ln\hat{k}-\ln\hat{k}^*) \\ & + \big[\alpha s_k\tilde{A}\cdot e^{-(1-\alpha)\ln k^*}\cdot e^{\hat{\eta}\ln h^*}\cdot\eta \\ & + \eta s_h\tilde{A}\cdot e^{\alpha\ln\hat{k}}\cdot e^{-(1-\eta)\ln\hat{h}^*}\cdot[-(1-\eta)]\big]\cdot(\ln\hat{h}-\ln\hat{h}^*)\end{aligned}$$

The steady-state conditions derived from equations (1.55) and (1.56) can be used to get

$$\begin{aligned}\dot{\hat{y}}/\hat{y} &= -(1-\alpha-\eta)\cdot(\delta+n+x)\cdot[\alpha\cdot(\ln\hat{k}-\ln\hat{k}^*)+\eta\cdot(\ln\hat{h}-\ln\hat{h}^*)] \\ &= -\beta^*\cdot(\ln\hat{y}-\ln\hat{y}^*)\end{aligned} \tag{1.57}$$

Therefore, in the neighborhood of the steady state, the convergence coefficient is $\beta^* = (1-\alpha-\eta)\cdot(\delta+n+x)$, just as in equation (1.54).

1.3 Models of Endogenous Growth

1.3.1 Theoretical Dissatisfaction with Neoclassical Theory

In the mid-1980s it became increasingly clear that the standard neoclassical growth model was theoretically unsatisfactory as a tool to explore the determinants of long-run growth. We have seen that the model without technological change predicts that the economy will eventually converge to a steady state with zero per capita growth. The fundamental reason is the diminishing returns to capital. One way out of this problem was to broaden the concept of capital, notably to include human components, and then assume that diminishing returns did not apply to this broader class of capital. This approach is the one outlined in the next section and explored in detail in chapters 4 and 5. However, another view was that technological progress in the form of the generation of new ideas was the only way that an economy could escape from diminishing returns in the long run. Thus it became a priority to go beyond the treatment of technological progress as exogenous and, instead, to explain this progress within the model of growth. However, endogenous approaches to technological change encountered basic problems within the neoclassical model—the essential reason is the nonrival nature of the ideas that underlie technology.

Remember that a key characteristic of the state of technology, T, is that it is a nonrival input to the production process. Hence, the replication argument that we used before to justify the assumption of constant returns to scale suggests that the correct measure of scale is the two rival inputs, capital and labor. Hence, the concept of constant returns to scale that we used is homogeneity of degree one in K and L:

$$F(\lambda K, \lambda L, T) = \lambda \cdot F(K, L, T)$$

Recall also that Euler's theorem implies that a function that is homogeneous of degree one can be decomposed as

$$F(K, L, T) = F_K \cdot K + F_L \cdot L \tag{1.58}$$

In our analysis up to this point, we have been assuming that the same technology, T, is freely available to all firms. This availability is technically feasible because T is nonrival. However, it may be that T is at least partly excludable—for example, patent protection, secrecy, and experience might allow some producers to have access to technologies that are superior to those available to others. For the moment, we maintain the assumption that technology is nonexcludable, so that all producers have the same access. This assumption also means that a technological advance is immediately available to all producers.

We know from our previous analysis that perfectly competitive firms that take the input prices, R and w, as given end up equating the marginal products to the respective input prices, that is, $F_K = R$ and $F_L = w$. It follows from equation (1.58) that the factor payments exhaust the output, so that each firm's profit equals zero at every point in time.

Suppose that a firm has the option to pay a fixed cost, κ, to improve the technology from T to T'. Since the new technology would, by assumption, be freely available to all other producers, we know that the equilibrium values of R and w would again entail a zero flow of profit for each firm. Therefore, the firm that paid the fixed cost, κ, will end up losing money overall, because the fixed cost would not be recouped by positive profits at any future dates. It follows that the competitive, neoclassical model cannot sustain purposeful investment in technical change if technology is nonexcludable (as well as nonrival).

The obvious next step is to allow the technology to be at least partly excludable. To bring out the problems with this extension, consider the polar case of full excludability, that is, where each firm's technology is completely private. Assume, however, that there are infinitely many ways in which firms can improve knowledge from T to T' by paying the fixed cost κ—in other words, there is free entry into the business of creating formulas. Suppose

If we substitute $f(k)/k = A$ in equation (1.13), we get

$$\dot{k}/k = sA - (n + \delta)$$

We return here to the case of zero technological progress, $x = 0$, because we want to show that per capita growth can now occur in the long run even without exogenous technological change. For a graphical presentation, the main difference is that the downward-sloping saving curve, $s \cdot f(k)/k$, in figure 1.4 is replaced in figure 1.12 by the horizontal line at the level sA. The depreciation curve is still the same horizontal line at $n + \delta$. Hence, $\dot{k}/k$ is the vertical distance between the two lines, sA and $n + \delta$. We depict the case in which $sA > (n + \delta)$, so that $\dot{k}/k > 0$. Since the two lines are parallel, $\dot{k}/k$ is constant; in particular, it is independent of k. Therefore, k always grows at the steady-state rate, $(\dot{k}/k)^* = sA - (n + \delta)$.

Since $y = Ak$, $\dot{y}/y = \dot{k}/k$ at every point in time. In addition, since $c = (1 - s) \cdot y$, $\dot{c}/c = \dot{k}/k$ also applies. Hence, all the per capita variables in the model always grow at the same, constant rate, given by

$$\gamma^* = sA - (n + \delta) \tag{1.60}$$

Note that an economy described by the AK technology can display positive long-run per capita growth without any technological progress. Moreover, the per capita growth rate

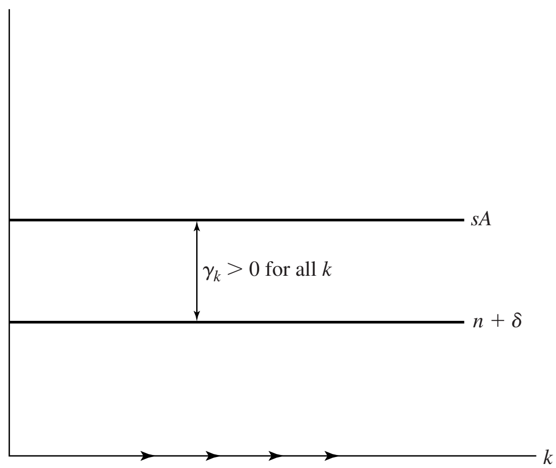

Figure 1.12
The *AK* Model. If the technology is AK, the saving curve, $s \cdot f(k)/k$, is a horizontal line at the level sA. If $sA > n + \delta$, perpetual growth of k occurs, even without technological progress.

that all firms begin with the technology T. Would an individual firm then have the incentive to pay κ to improve the technology to T'? In fact, the incentive appears to be enormous. At the existing input prices, R and w, a neoclassical firm with a superior technology would make a pure profit on each unit produced. Because of the assumed constant returns to scale, the firm would be motivated to hire all the capital and labor available in the economy. In this case, the firm would have lots of monopoly power and would likely no longer act as a perfect competitor in the goods and factor markets. So, the assumptions of the competitive model would break down.

A more basic problem with this result is that other firms would have perceived the same profit opportunity and would also have paid the cost κ to acquire the better technology, T'. However, when many firms improve their technology by the same amount, the competition pushes up the factor prices, R and w, so that the flow of profit is again zero. In this case, none of the firms can cover their fixed cost, κ, just as in the model in which technology was nonexcludable. Therefore, it is not an equilibrium for technological advance to occur (because all innovators make losses) and it is also not an equilibrium for this advance not to occur (because the potential profit to a single innovator is enormous).

These conceptual difficulties motivated researchers to introduce some aspects of imperfect competition to construct satisfactory models in which the level of the technology can be advanced by purposeful activity, such as R&D expenditures. This potential for endogenous technological progress and, hence, *endogenous growth,* may allow an escape from diminishing returns at the aggregate level. Models of this type were pioneered by Romer (1990) and Aghion and Howitt (1992); we consider them in chapters 6–8. For now, we deal only with models in which technology is either fixed or varying in an exogenous manner.

1.3.2 The *AK* Model

The key property of this class of endogenous-growth models is the absence of diminishing returns to capital. The simplest version of a production function without diminishing returns is the AK function:[29]

$$Y = AK \tag{1.59}$$

where A is a positive constant that reflects the level of the technology. The global absence of diminishing returns may seem unrealistic, but the idea becomes more plausible if we think of K in a broad sense to include human capital.[30] Output per capita is $y = Ak$, and the average and marginal products of capital are constant at the level $A > 0$.

29. We think that the first economist to use a production function of the AK type was von Neumann (1937).

30. Knight (1944) stressed the idea that diminishing returns might not apply to a broad concept of capital.

shown in equation (1.60) depends on the behavioral parameters of the model, including s, A, and n. For example, unlike the neoclassical model, a higher saving rate, s, leads to a higher rate of long-run per capita growth, γ^*.[31] Similarly if the level of the technology, A, improves once and for all (or if the elimination of a governmental distortion effectively raises A), then the long-run growth rate is higher. Changes in the rates of depreciation, δ, and population growth, n, also have permanent effects on the per capita growth rate.

Unlike the neoclassical model, the AK formulation does not predict absolute or conditional convergence, that is, $\partial(\dot{y}/y)/\partial y = 0$ applies for all levels of y. Consider a group of economies that are structurally similar in that the parameters s, A, n, and δ are the same. The economies differ only in terms of their initial capital stocks per person, $k(0)$, and, hence, in $y(0)$ and $c(0)$. Since the model says that each economy grows at the same per capita rate, γ^*, regardless of its initial position, the prediction is that all the economies grow at the same per capita rate. This conclusion reflects the absence of diminishing returns. Another way to see this result is to observe that the AK model is just a Cobb–Douglas model with a unit capital share, $\alpha = 1$. The analysis of convergence in the previous section showed that the speed of convergence was given in equation (1.45) by $\beta^* = (1 - \alpha) \cdot (x + n + \delta)$; hence, $\alpha = 1$ implies $\beta^* = 0$. This prediction is a substantial failing of the model, because conditional convergence appears to be an empirical regularity. See chapters 11 and 12 for a detailed discussion.

We mentioned that one way to think about the absence of diminishing returns to capital in the AK production function is to consider a broad concept of capital that encompassed physical and human components. In chapters 4 and 5 we consider in more detail models that allow for these two types of capital.

Other approaches have been used to eliminate the tendency for diminishing returns in the neoclassical model. We study in chapter 4 the notion of learning by doing, which was introduced by Arrow (1962) and used by Romer (1986). In these models, the experience with production or investment contributes to productivity. Moreover, the learning by one producer may raise the productivity of others through a process of spillovers of knowledge from one producer to another. Therefore, a larger economy-wide capital stock (or a greater cumulation of the aggregate of past production) improves the level of the technology for each producer. Consequently, diminishing returns to capital may not apply in the aggregate, and increasing returns are even possible. In a situation of increasing returns, each producer's average

31. With the AK production function, we can never get the kind of inefficient oversaving that is possible in the neoclassical model. A shift at some point in time to a permanently higher s means a lower level of c at that point but a permanently higher per capita growth rate, γ^*, and, hence, higher levels of c after some future date. This change cannot be described as inefficient because it may be desirable or undesirable depending on how households discount future levels of consumption.

product of capital, $f(k)/k$, tends to rise with the economy-wide value of k. Consequently, the $s \cdot f(k)/k$ curve in figure 1.4 tends to be upward sloping, at least over some range, and the growth rate, $\dot{k}/k$, rises with k in this range. Thus these kinds of models predict at least some intervals of per capita income in which economies tend to diverge. It is unclear, however, whether these divergence intervals are present in the data.

1.3.3 Endogenous Growth with Transitional Dynamics

The AK model delivers endogenous growth by avoiding diminishing returns to capital in the long run. This particular production function also implies, however, that the marginal and average products of capital are always constant and, hence, that growth rates do not exhibit the convergence property. It is possible to retain the feature of constant returns to capital in the long run, while restoring the convergence property—an idea brought out by Jones and Manuelli (1990).[32]

Consider again the expression for the growth rate of k from equation (1.13):

$$\dot{k}/k = s \cdot f(k)/k - (n + \delta) \tag{1.61}$$

If a steady state exists, the associated growth rate, $(\dot{k}/k)^*$, is constant by definition. A positive $(\dot{k}/k)^*$ means that k grows without bound. Equation (1.13) implies that it is necessary and sufficient for $(\dot{k}/k)^*$ to be positive to have the average product of capital, $f(k)/k$, remain above $(n + \delta)/s$ as k approaches infinity. In other words, if the average product approaches some limit, then $\lim_{k\to\infty}[f(k)/k] > (n + \delta)/s$ is necessary and sufficient for endogenous, steady-state growth.

If $f(k) \to \infty$ as $k \to \infty$, then an application of l'Hôpital's rule shows that the limits as k approaches infinity of the average product, $f(k)/k$, and the marginal product, $f'(k)$, are the same. (We assume here that $\lim_{k\to\infty}[f'(k)]$ exists.) Hence, the key condition for endogenous, steady-state growth is that $f'(k)$ be bounded sufficiently far above 0:

$$\lim_{k\to\infty}[f(k)/k] = \lim_{k\to\infty}[f'(k)] > (n + \delta)/s > 0$$

This inequality violates one of the standard Inada conditions in the neoclassical model, $\lim_{k\to\infty}[f'(k)] = 0$. Economically, the violation of this condition means that the tendency for diminishing returns to capital tends to disappear. In other words, the production function can exhibit diminishing or increasing returns to k when k is low, but the marginal product of capital must be bounded from below as k becomes large. A simple example, in which the production function converges asymptotically to the AK form, is

$$Y = F(K, L) = AK + BK^{\alpha}L^{1-\alpha} \tag{1.62}$$

32. See Kurz (1968) for a related discussion.

where $A > 0$, $B > 0$, and $0 < \alpha < 1$. Note that this production function is a combination of the AK and Cobb–Douglas functions. It exhibits constant returns to scale and positive and diminishing returns to labor and capital. However, one of the Inada conditions is violated because $\lim_{K\to\infty}(F_K) = A > 0$.

We can write the function in per capita terms as

$$y = f(k) = Ak + Bk^{\alpha}$$

The average product of capital is given by

$$f(k)/k = A + Bk^{-(1-\alpha)}$$

which is decreasing in k but approaches A as k tends to infinity.

The dynamics of this model can be analyzed with the usual expression from equation (1.13):

$$\dot{k}/k = s \cdot \left[A + Bk^{-(1-\alpha)}\right] - (n + \delta) \tag{1.63}$$

Figure 1.13 shows that the saving curve is downward sloping, and the line $n + \delta$ is horizontal. The difference from figure 1.4 is that, as k goes to infinity, the saving curve in figure 1.13 approaches the positive quantity sA, rather than 0. If $sA > n + \delta$, as assumed in the figure, the steady-state growth rate, $(\dot{k}/k)^{*}$, is positive.

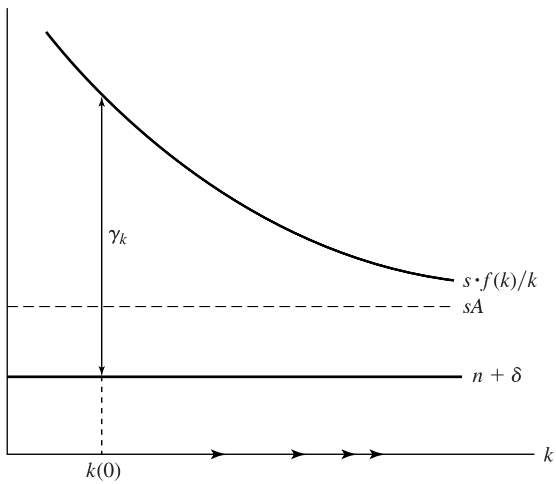

Figure 1.13
Endogenous growth with transitional dynamics. If the technology is $F(K, L) = AK + BK^{\alpha}L^{1-\alpha}$, the growth rate of k is diminishing for all k. If $sA > n + \delta$, the growth rate of k asymptotically approaches a positive constant, given by $sA - n - \delta$. Hence, endogenous growth coexists with a transition in which the growth rate diminishes as the economy develops.

This model yields endogenous, steady-state growth but also predicts conditional convergence, as in the neoclassical model. The reason is that the convergence property derives from the inverse relation between $f(k)/k$ and k, a relation that still holds in the model. Figure 1.13 shows that if two economies differ only in terms of their initial values, $k(0)$, the one with the smaller capital stock per person will grow faster in per capita terms.

1.3.4 Constant-Elasticity-of-Substitution Production Functions

Consider as another example the production function (due to Arrow et al., 1961) that has a constant elasticity of substitution (CES) between labor and capital:

$$Y = F(K, L) = A \cdot \{a \cdot (bK)^{\psi} + (1-a) \cdot [(1-b) \cdot L]^{\psi}\}^{1/\psi} \tag{1.64}$$

where $0 < a < 1$, $0 < b < 1$,[33] and $\psi < 1$. Note that the production function exhibits constant returns to scale for all values of ψ. The elasticity of substitution between capital and labor is $1/(1-\psi)$ (see the appendix, section 1.5.4). As $\psi \to -\infty$, the production function approaches a fixed-proportions technology (discussed in the next section), $Y = \min[bK, (1-b)L]$, where the elasticity of substitution is 0. As $\psi \to 0$, the production function approaches the Cobb–Douglas form, $Y = (\text{constant}) \cdot K^a L^{1-a}$, and the elasticity of substitution is 1 (see the appendix, section 1.5.4). For $\psi = 1$, the production function is linear, $Y = A \cdot [abK + (1-a) \cdot (1-b) \cdot L]$, so that K and L are perfect substitutes (infinite elasticity of substitution).

Divide both sides of equation (1.64) by L to get an expression for output per capita:

$$y = f(k) = A \cdot [a \cdot (bk)^{\psi} + (1-a) \cdot (1-b)^{\psi}]^{1/\psi}$$

The marginal and average products of capital are given, respectively, by

$$f'(k) = Aab^{\psi}[ab^{\psi} + (1-a) \cdot (1-b)^{\psi} \cdot k^{-\psi}]^{(1-\psi)/\psi}$$

$$f(k)/k = A[ab^{\psi} + (1-a) \cdot (1-b)^{\psi} \cdot k^{-\psi}]^{1/\psi}$$

Thus, $f'(k)$ and $f(k)/k$ are each positive and diminishing in k for all values of ψ.

We can study the dynamic behavior of a CES economy by returning to the expression from equation (1.13):

$$\dot{k}/k = s \cdot f(k)/k - (n + \delta) \tag{1.65}$$

33. The standard formulation does not include the terms b and $1-b$. The implication then is that the shares of K and L in total product each approach one-half as $\psi \to -\infty$. In our formulation, the shares of K and L approach b and $1-b$, respectively, as $\psi \to -\infty$.

If we graph versus k, then $s \cdot f(k)/k$ is a downward-sloping curve, $n + \delta$ is a horizontal line, and $\dot{k}/k$ is still represented by the vertical distance between the curve and the line. The behavior of the growth rate now depends, however, on the parameter ψ, which governs the elasticity of substitution between L and K.

Consider first the case $0 < \psi < 1$, that is, a high degree of substitution between L and K. The limits of the marginal and average products of capital in this case are

$$\lim_{k\to\infty}[f'(k)] = \lim_{k\to\infty}[f(k)/k] = Aba^{1/\psi} > 0$$
$$\lim_{k\to 0}[f'(k)] = \lim_{k\to 0}[f(k)/k] = \infty$$

Hence, the marginal and average products approach a positive constant, rather than 0, as k goes to infinity. In this sense, the CES production function with high substitution between the factors $(0 < \psi < 1)$ looks like the example in equation (1.62) in which diminishing returns vanished asymptotically. We therefore anticipate that this CES model can generate endogenous, steady-state growth.

Figure 1.14 shows the results graphically. The $s \cdot f(k)/k$ curve is downward sloping, and it asymptotes to the positive constant $sAb \cdot a^{1/\psi}$. If the saving rate is high enough, so that $sAb \cdot a^{1/\psi} > n + \delta$—as assumed in the figure—then the $s \cdot f(k)/k$ curve always lies above the $n + \delta$ line. In this case, the per capita growth rate is always positive, and the model

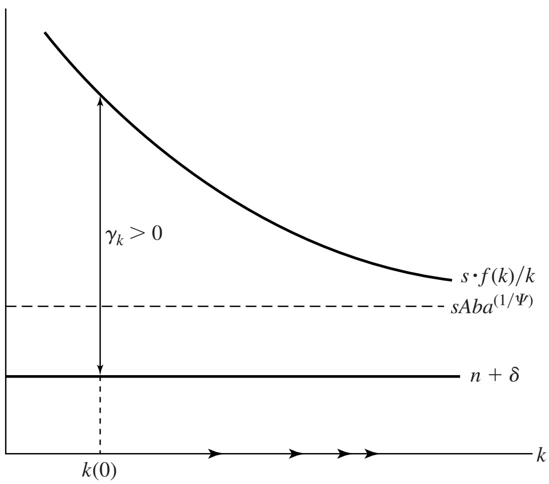

Figure 1.14
The CES model with $0 < \psi < 1$ and $sAb \cdot a^{1/\psi} > n + \delta$. If the CES technology exhibits a high elasticity of substitution $(0 < \psi < 1)$, endogenous growth arises if the parameters satisfy the inequality $sAb \cdot a^{1/\psi} > n + \delta$. Along the transition, the growth rate of k diminishes.

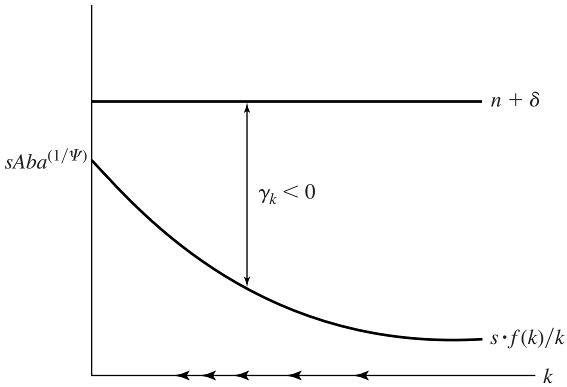

Figure 1.15
The CES model with $\psi < 0$ and $sAb \cdot a^{1/\psi} < n + \delta$. If the CES technology exhibits a low elasticity of substitution ($\psi < 0$), the growth rate of k would be negative for all levels of k if $sAb \cdot a^{1/\psi} < n + \delta$.

generates endogenous, steady-state growth at the rate

$$\gamma^* = sAb \cdot a^{1/\psi} - (n + \delta)$$

The dynamics of this model are similar to those described in figure 1.13.[34]

Assume now $\psi < 0$, that is, a low degree of substitution between L and K. The limits of the marginal and average products of capital in this case are

$$\lim_{k\to\infty} [f'(k)] = \lim_{k\to\infty} [f(k)/k] = 0$$
$$\lim_{k\to 0} [f'(k)] = \lim_{k\to 0} [f(k)/k] = Ab \cdot a^{1/\psi} < \infty$$

Since the marginal and average products approach 0 as k approaches infinity, the key Inada condition is satisfied, and the model does not generate endogenous growth. In this case, however, the violation of the Inada condition as k approaches 0 may cause problems. Suppose that the saving rate is low enough so that $sAb \cdot a^{1/\psi} < n + \delta$. In this case, the $s \cdot f(k)/k$ curve starts at a point below $n + \delta$, and it converges to 0 as k approaches infinity. Figure 1.15 shows, accordingly, that the curve never crosses the $n + \delta$ line, and, hence, no steady state exists with a positive value of k. Since the growth rate $\dot{k}/k$ is always negative, the economy shrinks over time, and k, y, and c all approach 0.[35]

34. If $0 < \psi < 1$ and $sAb \cdot a^{1/\psi} < n + \delta$, then the $s \cdot f(k)/k$ curve crosses $n + \delta$ at the steady-state value k^*, as in the standard neoclassical model of figure 1.4. Endogenous growth does not apply in this case.

35. If $\psi < 0$ and $sAb \cdot a^{1/\psi} > n + \delta$, then the $s \cdot f(k)/k$ curve again intersects the $n + \delta$ line at the steady-state value k^*.

Since the average product of capital, $f(k)/k$, is a negative function of k for all values of ψ, the growth rate $\dot{k}/k$ is also a negative function of k. The CES model therefore always exhibits the convergence property: for two economies with identical parameters and different initial values, $k(0)$, the one with the lower value of $k(0)$ has the higher value of $\dot{k}/k$. When the parameters differ across economies, the model predicts conditional convergence, as described before.

We can use the method developed earlier for the case of a Cobb–Douglas production function to derive a formula for the convergence coefficient in the neighborhood of the steady state. The result for a CES production function, which extends equation (1.45), is[36]

$$\beta^* = -(x+n+\delta) \cdot \left[1 - a \cdot \left(\frac{bsA}{x+n+\delta}\right)^{\psi}\right] \tag{1.66}$$

For the Cobb–Douglas case, where $\psi = 0$ and $a = \alpha$, equation (1.66) reduces to equation (1.45). For $\psi \neq 0$, a new result is that β^* in equation (1.66) depends on s and A. If $\psi > 0$ (high substitutability between L and K), then β^* falls with sA, and vice versa if $\psi < 0$. The coefficient β^* is independent of s and A only in the Cobb–Douglas case, where $\psi = 0$.

1.4 Other Production Functions . . . Other Growth Theories

1.4.1 The Leontief Production Function and the Harrod–Domar Controversy

A production function that was used prior to the neoclassical one is the Leontief (1941), or fixed-proportions, function,

$$Y = F(K, L) = \min(AK, BL) \tag{1.67}$$

where $A > 0$ and $B > 0$ are constants. This specification, which corresponds to $\psi \to -\infty$ in the CES form in equation (1.64), was used by Harrod (1939) and Domar (1946). With fixed proportions, if the available capital stock and labor force happen to be such that $AK = BL$, then all workers and machines are fully employed. If K and L are such that $AK > BL$, then only the quantity of capital $(B/A) \cdot L$ is used, and the remainder remains idle. Conversely, if $AK < BL$, then only the amount of labor $(A/B) \cdot K$ is used, and the remainder is unemployed. The assumption of no substitution between capital and labor led Harrod and Domar to predict that capitalist economies would have undesirable outcomes in the form of perpetual increases in unemployed workers or machines. We provide here a brief analysis of the Harrod–Domar model using the tools developed earlier in this chapter.

36. See Chua (1993) for additional discussion. The formula for β in equation (1.66) applies only for cases in which the steady-state level k^* exists. If $0 < \psi < 1$, it applies for $bsA \cdot a^{1/\psi} < x+n+\delta$. If $\psi < 0$, it applies for $bsA \cdot a^{1/\psi} > x+n+\delta$.

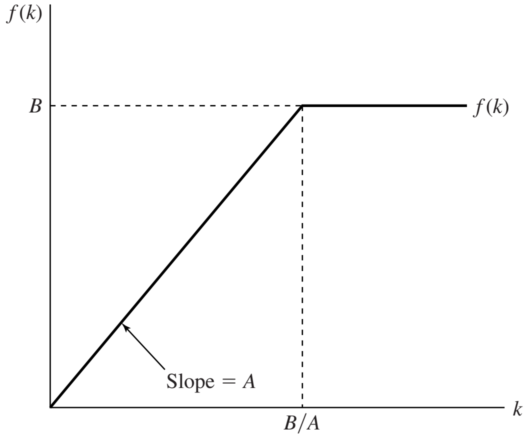

Figure 1.16
The Leontief production function in per capita terms. In per capita terms, the Leontief production function can be written as $y = \min(Ak, B)$. For $k < B/A$, output per capita is given by $y = Ak$. For $k > B/A$, output per capita is given by $y = B$.

Divide both sides of equation (1.67) by L to get output per capita:

$$y = \min(Ak, B)$$

For $k < B/A$, capital is fully employed, and $y = Ak$. Hence, figure 1.16 shows that the production function in this range is a straight line from the origin with slope A. For $k > B/A$, the quantity of capital used is constant, and Y is the constant multiple B of labor, L. Hence, output per worker, y, equals the constant B, as shown by the horizontal part of $f(k)$ in the figure. Note that, as k approaches infinity, the marginal product of capital, $f'(k)$, is zero. Hence, the key Inada condition is satisfied, and we do not expect this production function to yield endogenous steady-state growth.

We can use the expression from equation (1.13) to get

$$\dot{k}/k = s \cdot [\min(Ak, B)]/k - (n + \delta) \tag{1.68}$$

Figures 1.17*a* and 1.17*b* show that the first term, $s \cdot [\min(Ak, B)]/k$, is a horizontal line at sA for $k \leq B/A$. For $k > B/A$, this term is a downward-sloping curve that approaches zero as k goes to infinity. The second term in equation (1.68) is the usual horizontal line at $n + \delta$.

Assume first that the saving rate is low enough so that $sA < n + \delta$, as depicted in figure 1.17. The saving curve, $s \cdot f(k)/k$, then never crosses the $n + \delta$ line, so there is no

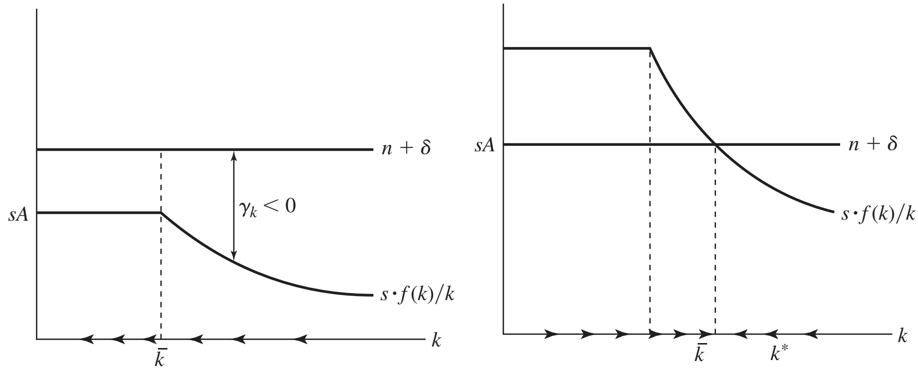

Figure 1.17
The Harrod–Domar model. In panel *a*, which assumes $sA < n + \delta$, the growth rate of k is negative for all k. Therefore, the economy approaches $k = 0$. In panel *b*, which assumes $sA > n + \delta$, the growth rate of k is positive for $k < k^*$ and negative for $k > k^*$, where k^* is the stable steady-state value. Since k^* exceeds B/A, a part of the capital stock always remains idle. Moreover, the quantity of idle capital grows steadily (along with K and L).

positive steady-state value, k^*. Moreover, the growth rate of capital, $\dot{k}/k$, is always negative, so the economy shrinks in per capita terms, and k, y, and c all approach 0. The economy therefore ends up to the left of B/A and has permanent and increasing unemployment.

Suppose now that the saving rate is high enough so that $sA > n + \delta$, as shown in figure 1.17b. Since the $s \cdot f(k)/k$ curve approaches 0 as k tends to infinity, this curve eventually crosses the $n + \delta$ line at the point $k^* > B/A$. Therefore, if the economy begins at $k(0) < k^*$, $\dot{k}/k$ equals the constant $sA - n - \delta > 0$ until k attains the value B/A. At that point, $\dot{k}/k$ falls until it reaches 0 at $k = k^*$. If the economy starts at $k(0) > k^*$, $\dot{k}/k$ is initially negative and approaches 0 as k approaches k^*.

Since $k^* > B/A$, the steady state features idle machines but no unemployed workers. Since k is constant in the steady state, the quantity K grows along with L at the rate n. Since the fraction of machines that are employed remains constant, the quantity of idle machines also grows at the rate n (yet households are nevertheless assumed to keep saving at the rate s).

The only way to reach a steady state in which all capital and labor are employed is for the parameters of the model to satisfy the condition $sA = n + \delta$. Since the four parameters that appear in this condition are all exogenous, there is no reason for the equality to hold. Hence, the conclusion from Harrod and Domar was that an economy would, in all probability, reach one of two undesirable outcomes: perpetual growth of unemployment or perpetual growth of idle machinery.

We know now that there are several implausible assumptions in the arguments of Harrod and Domar. First, the Solow–Swan model showed that Harrod and Domar's parameter A—the average product of capital—would typically depend on k, and k would adjust to satisfy the equality $s \cdot f(k)/k = n + \delta$ in the steady state. Second, the saving rate could adjust to satisfy this condition. In particular, if agents maximize utility (as we assume in the next chapter), they would not find it optimal to continue to save at the constant rate s when the marginal product of capital was zero. This adjustment of the saving rate would rule out an equilibrium with permanently idle machinery.

1.4.2 Growth Models with Poverty Traps

One theme in the literature of economic development concerns *poverty traps*.[37] We can think of a poverty trap as a stable steady state with low levels of per capita output and capital stock. This outcome is a trap because, if agents attempt to break out of it, the economy has a tendency to return to the low-level, stable steady state.

We observed that the average product of capital, $f(k)/k$, declines with k in the neoclassical model. We also noted, however, that this average product may rise with k in some models that feature increasing returns, for example, in formulations that involve learning by doing and spillovers. One way for a poverty trap to arise is for the economy to have an interval of diminishing average product of capital followed by a range of rising average product. (Poverty traps also arise in some models with nonconstant saving rates; see Galor and Ryder, 1989.)

We can get a range of increasing returns by imagining that a country has access to a traditional, as well as a modern, technology.[38] Imagine that producers can use a primitive production function, which takes the usual Cobb–Douglas form,

$$Y_A = AK^{\alpha}L^{1-\alpha} \tag{1.69}$$

The country also has access to a modern, higher productivity technology,[39]

$$Y_B = BK^{\alpha}L^{1-\alpha} \tag{1.70}$$

where $B > A$. However, in order to exploit this better technology, the country as a whole is assumed to have to pay a setup cost at every moment in time, perhaps to cover the necessary public infrastructure or legal system. We assume that this cost is proportional to

37. See especially the *big-push* model of Lewis (1954). A more modern formulation of this idea appears in Murphy, Shleifer, and Vishny (1989).

38. This section is an adaptation of Galor and Zeira (1993), who use two technologies in the context of education.

39. More generally, the capital intensity for the advanced technology would differ from that for the primitive technology. However, this extension complicates the algebra without making any substantive differences.

the labor force and given by bL, where $b > 0$. We assume further that this cost is borne by the government and financed by a tax at rate b on each worker. The results are the same whether the tax is paid by producers or workers (who are, in any event, the same persons an economy with household-producers).

In per worker terms, the first production function is

$$y_A = Ak^{\alpha} \tag{1.71}$$

The second production function, when considered net of the setup cost and in per worker terms, is

$$y_B = Bk^{\alpha} - b \tag{1.72}$$

The two production functions are drawn in figure 1.18.

If the government has decided to pay the setup cost, which equals b per worker, all producers will use the modern technology (because the tax b for each worker must be paid in any case). If the government has not paid the setup cost, all producers must use the primitive technology. A sensible government would pay the setup cost if the shift to the modern technology leads to an increase in output per worker at the existing value of k and when measured net of the setup cost. In the present setting, the shift is warranted if k exceeds a critical level, given by $\tilde{k} = [b/(B - A)]^{1/\alpha}$. Thus, the critical value of k rises with the setup cost parameter, b, and falls with the difference in the productivity parameters, $B - A$. We assume that the government pays the setup cost if $k \geq \tilde{k}$ and does not pay it if $k < \tilde{k}$.

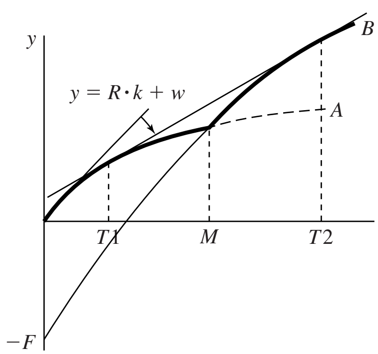

Figure 1.18
Traditional and modern production functions. The traditional production function has relatively low productivity. The modern production function exhibits higher productivity but is assumed to require a fixed cost to operate.

The growth rate of capital per worker is still given by the fundamental equation of the Solow–Swan model, equation (1.23), as

$$\dot{k}/k = s \cdot f(k)/k - (\delta + n)$$

where $f(k) = Ak^{\alpha}$ if $k < \tilde{k}$ and $f(k) = Bk^{\alpha} - b$ if $k \geq \tilde{k}$. The average product of capital, $f(k)/k$, can be measured graphically in figure 1.18 by the slope of the cord that goes from the origin to the effective production function. We can see that there is a range of $k \geq \tilde{k}$ where the average product is increasing. The saving curve therefore looks like the one depicted in figure 1.19: it has the familiar negative slope at low levels of k, is then followed by a range with a positive slope, and again has a negative slope at very high levels of k.

Figure 1.19 shows that the $s \cdot f(k)/k$ curve first crosses the $n + \delta$ line at the low steady-state value, k^*_{low}, where we assume here that $k^*_{low} < \tilde{k}$. This steady state has the properties that are familiar from the neoclassical model. In particular, $\dot{k}/k > 0$ for $k < k^*_{low}$, and $\dot{k}/k < 0$ at least in an interval of $k > k^*_{low}$. Hence, k^*_{low} is a stable steady state: it is a poverty trap in the sense described before.

The tendency for increasing returns in the middle range of k is assumed to be strong enough so that the $s \cdot f(k)/k$ curve eventually rises to cross the $n + \delta$ line again at the

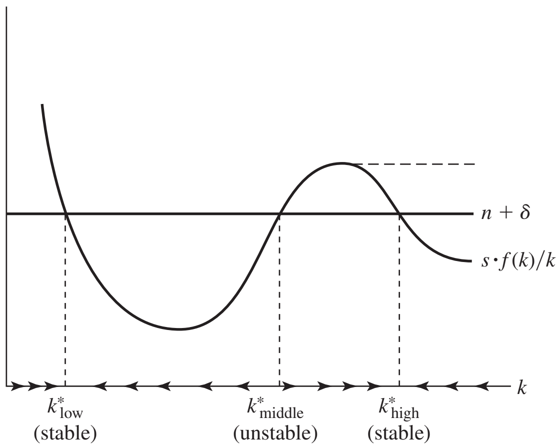

Figure 1.19
A poverty trap. The production function is assumed to exhibit diminishing returns to k when k is low, increasing returns for a middle range of k, and either constant or diminishing returns when k is high. The curve $s \cdot f(k)/k$ is therefore downward sloping for low values of k, upward sloping for an intermediate range of k, and downward sloping or horizontal for high values of k. The steady-state value k^*_{low} is stable and therefore constitutes a poverty trap for countries that begin with k between 0 and k^*_{middle}. If a country begins with $k > k^*_{middle}$, it converges to k^*_{high} if diminishing returns to k ultimately set in. If the returns to capital are constant at high values of k, as depicted by the dashed portion of the curve, the country converges to a positive long-run growth rate of k.

steady-state value k^*_{middle}. This steady state is, however, unstable, because $\dot{k}/k < 0$ applies to the left, and $\dot{k}/k > 0$ holds to the right. Thus, if the economy begins with $k^*_{\text{low}} < k(0) < k^*_{\text{middle}}$, its natural tendency is to return to the development trap at k^*_{low}, whereas if it manages somehow to get to $k(0) > k^*_{\text{middle}}$, it tends to grow further to reach still higher levels of k.

In the range where $k > k^*_{\text{middle}}$, the economy's tendency toward diminishing returns eventually brings $s \cdot f(k)/k$ down enough to equal $n + \delta$ at the steady-state value k^*_{high}. This steady state, corresponding to a high level of per capita income but to zero long-term per capita growth, is familiar from our study of the neoclassical model. The key problem for a less-developed economy at the trap level k^*_{low} is to get over the hump and thereby attain a high long-run level of per capita income.

One empirical implication of the model described by figure 1.19 is that there would exist a middle range of values of k—around k^*_{middle}—for which the growth rate, $\dot{k}/k$, is increasing in k and, hence, in y. That is, a divergence pattern should hold over this range of per capita incomes. Our reading of the evidence across countries, discussed in chapter 12, does not support this hypothesis. These results are, however, controversial—see, for example, Quah (1996).

1.5 Appendix: Proofs of Various Propositions

1.5.1 Proof That Each Input Is Essential for Production with a Neoclassical Production Function

We noted in the main body of this chapter that the neoclassical properties of the production function imply that the two inputs, K and L, are each essential for production. To verify this proposition, note first that if $Y \to \infty$ as $K \to \infty$, then

$$\lim_{K\to\infty} \frac{Y}{K} = \lim_{K\to\infty} \frac{\partial Y}{\partial K} = 0$$

where the first equality comes from l'Hôpital's rule and the second from the Inada condition. If Y remains bounded as K tends to infinity, then

$$\lim_{K\to\infty} (Y/K) = 0$$

follows immediately. We also know from constant returns to scale that, for any finite L,

$$\lim_{K\to\infty} (Y/K) = \lim_{K\to\infty} [F(1, L/K)] = F(1, 0)$$

so that $F(1, 0) = 0$. The condition of constant returns to scale then implies

$$F(K, 0) = K \cdot F(1, 0) = 0$$

for any finite K. We can show from an analogous argument that $F(0, L) = 0$ for any finite L. These results verify that each input is essential for production.

To demonstrate that output goes to infinity when either input goes to infinity, note that

$$F(K, L) = L \cdot f(k) = K \cdot [f(k)/k]$$

Therefore, for any finite K,

$$\lim_{L\to\infty}[F(K, L)] = K \cdot \lim_{k\to 0}[f(k)/k] = K \cdot \lim_{k\to 0}[f'(k)] = \infty$$

where the last equalities follow from l'Hôpital's rule (because essentiality implies $f[0] = 0$) and the Inada condition. We can show from an analogous argument that $\lim_{K\to\infty}[F(K, L)] = \infty$. Therefore, output goes to infinity when either input goes to infinity.

1.5.2 Properties of the Convergence Coefficient in the Solow–Swan Model

Equation (1.46) is a log-linearization of equation (1.41) around the steady-state position. To obtain equation (1.46), we have to rewrite equation (1.41) in terms of $\log(\hat{k})$. Note that $\dot{\hat{k}}/\hat{k}$ is the time derivative of $\log(\hat{k})$, and $(\hat{k})^{-(1-\alpha)}$ can be written as $e^{-(1-\alpha)\cdot\log(\hat{k})}$. The steady-state value of $sA(\hat{k})^{-(1-\alpha)}$ equals $x + n + \delta$. We can now take a first-order Taylor expansion of $\log(\hat{k})$ around $\log(\hat{k}^*)$ to get equation (1.46). See the appendix on mathematics at the end of the book for additional discussion. This result appears in Sala-i-Martin (1990) and Mankiw, Romer, and Weil (1992).

The true speed of convergence for $\hat{k}$ or $\hat{y}$ is not constant; it depends on the distance from the steady state. The growth rate of $\hat{y}$ can be written as

$$\dot{\hat{y}}/\hat{y} = \alpha \cdot \left[s \cdot A^{1/\alpha} \cdot (\hat{y})^{-(1-\alpha)/\alpha} - (x + n + \delta)\right]$$

If we use the condition $\hat{y}^* = A \cdot [sA/(x+n+\delta)]^{\alpha/(1-\alpha)}$, we can express the growth rate as

$$\dot{\hat{y}}/\hat{y} = \alpha \cdot (x + n + \delta) \cdot \left[(\hat{y}/\hat{y}^*)^{-(1-\alpha)/\alpha} - 1\right]$$

The convergence coefficient is

$$\beta = -d(\dot{\hat{y}}/\hat{y})]/d[\log(\hat{y})] = (1 - \alpha) \cdot (x + n + \delta) \cdot (\hat{y}/\hat{y}^*)^{-(1-\alpha)/\alpha}$$

At the steady state, $\hat{y} = \hat{y}^*$ and $\beta = (1 - \alpha) \cdot (x + n + \delta)$, as in equation (1.45). More generally, β declines as $\hat{y}/\hat{y}^*$ rises.

1.5.3 Proof That Technological Progress Must Be Labor Augmenting

We mentioned in the text that technological progress must take the labor-augmenting form shown in equation (1.34) in order for the model to have a steady state with constant growth rates. To prove this result, we start by assuming a production function that includes labor-augmenting and capital-augmenting technological progress:

$$Y = F[K \cdot B(t), L \cdot A(t)] \tag{1.73}$$

where $B(t) = A(t)$ implies that the technological progress is Hicks neutral.

We assume that $A(t) = e^{xt}$ and $B(t) = e^{zt}$, where $x \geq 0$ and $z \geq 0$ are constants. If we divide both sides of equation (1.73) by K, we can express output per unit of capital as

$$Y/K = e^{zt} \cdot \left\{ F\left[1, \frac{L \cdot A(t)}{K \cdot B(t)}\right] \right\} = e^{zt} \cdot \varphi\left[(L/K) \cdot e^{(x-z) \cdot t}\right]$$

where $\varphi(\cdot) \equiv F[1, \frac{L \cdot A(t)}{K \cdot B(t)}]$. The population, L, grows at the constant rate n. If γ_K^* is the constant growth rate of K in the steady state, the expression for Y/K can be written as

$$Y/K = e^{zt} \cdot \varphi\left[e^{(n+x-z-\gamma_K^*) \cdot t}\right] \tag{1.74}$$

Recall that the growth rate of K is given by

$$\dot{K}/K = s \cdot (Y/K) - \delta$$

In the steady state, $\dot{K}/K$ equals the constant γ_K^*, and, hence, Y/K must be constant. There are two ways to get the right-hand side of equation (1.74) to be constant. First, $z = 0$ and $\gamma_K^* = n + x$; that is, technological progress is solely labor augmenting, and the steady-state growth rate of capital equals $n + x$. In this case, the production function can be written in the form of equation (1.34).

The second way to get the right-hand side of equation (1.74) to be constant is with $z \neq 0$ and for the term $\varphi[e^{(n+x-z-\gamma_K^*)t}]$ exactly to offset the term e^{zt}. For this case to apply, the derivative of Y/K (in the proposed steady state) with respect to time must be identically zero. If we take the derivative of equation (1.74), set it to zero, and rearrange terms, we get

$$\varphi'(\chi) \cdot \chi / \varphi(\chi) = -z/(n + x - z - \gamma_K^*)$$

where $\chi \equiv e^{(n+x-z-\gamma_K^*) \cdot t}$, and the right-hand side is a constant. If we integrate out, we can write the solution as

$$\varphi(\chi) = (\text{constant}) \cdot \chi^{1-\alpha}$$

where α is a constant. This result implies that the production function can be written as

$$Y = (\text{constant}) \cdot (Ke^{zt})^{\alpha} \cdot (Le^{xt})^{1-\alpha} = (\text{constant}) \cdot K^{\alpha} \cdot (Le^{\nu t})^{1-\alpha}$$

where $\nu = [z\alpha + x \cdot (1 - \alpha)]/(1 - \alpha)$. In other words, if the rate of capital-augmenting technological progress, z, is nonzero and a steady state exists, the production function must take the Cobb–Douglas form. Moreover, if the production function is Cobb–Douglas, we can always express technological change as purely labor augmenting (at the rate ν). The conclusion, therefore, is that the existence of a steady state implies that technological progress can be written in the labor-augmenting form.

Another approach to technological progress assumes that capital goods produced later—that is, in a more recent *vintage*—are of higher quality for a given cost. If quality improves in accordance with $T(t)$, the equation for capital accumulation in this vintage model is

$$\dot{K} = s \cdot T(t) \cdot F(K, L) - \delta K \tag{1.75}$$

where K is measured in units of constant quality. This equation corresponds to Hicks-neutral technological progress given by $T(t)$ in the production function. The only difference from the standard specification is that output is $Y = F(K, L)$—not $T(t) \cdot F(K, L)$.

If we want to use a model that possesses a steady state, we would still have to assume that $F(K, L)$ was Cobb–Douglas. In that case, the main properties of the vintage model turn out to be indistinguishable from those of the model that we consider in the text in which technological progress is labor augmenting (see Phelps, 1962, and Solow, 1969, for further discussion). One difference in the vintage model is that, although K and Y grow at constant rates in the steady state, the growth rate of K (in units of constant quality) exceeds that of Y. Hence, K/Y is predicted to rise steadily in the long run.

1.5.4 Properties of the CES Production Function

The elasticity of substitution is a measure of the curvature of the isoquants. The slope of an isoquant is

$$\left.\frac{dL}{dK}\right|_{\text{isoquant}} = -\frac{\partial F(\cdot)/\partial K}{\partial F(\cdot)/\partial L}$$

The elasticity is given by

$$\left[\frac{\partial(\text{Slope})}{\partial(L/K)} \cdot \frac{L/K}{\text{Slope}}\right]^{-1}$$

For the CES production function shown in equation (1.64), the slope of the isoquant is

$$-(L/K)^{1-\psi} \cdot a \cdot b^{\psi}/[(1-a) \cdot (1-b)^{\psi}]$$

and the elasticity is $1/(1-\psi)$, a constant.

To compute the limit of the production function as ψ approaches 0, use equation (1.64) to get $\lim_{\psi \to 0}[\log(Y)] = \log(A) + 0/0$, which involves an indeterminate form. Apply l'Hôpital's rule to get

$$\begin{aligned}
&\lim_{\psi\to 0}[\log(Y)] \\
&\quad = \log(A) + \left[\frac{a(bK)^{\psi}\cdot\log(bK) + (1-a)\cdot[(1-b)\cdot L]^{\psi}\cdot\log[(1-b)\cdot L]}{a\cdot(bK)^{\psi} + (1-a)\cdot[(1-b)\cdot L]^{\psi}}\right]_{\psi=0} \\
&\quad = \log(A) + a\cdot\log(bK) + (1-a)\cdot\log[(1-b)\cdot L]
\end{aligned}$$

It follows that $Y = \tilde{A}K^{a}L^{1-a}$, where $\tilde{A} = Ab^{a}\cdot(1-b)^{1-a}$. That is, the CES production function approaches the Cobb–Douglas form as ψ tends to zero.

1.6 Problems

1.1 Convergence.

a. Explain the differences among absolute convergence, conditional convergence, and a reduction in the dispersion of real per capita income across groups.

b. Under what circumstances does absolute convergence imply a decline in the dispersion of per capita income?

1.2 Forms of technological progress. Assume that the rate of exogenous technological progress is constant.

a. Show that a steady state can coexist with technological progress only if this progress takes a labor-augmenting form. What is the intuition for this result?

b. Assume that the production function is $Y = F[B(T)\cdot K, A(t)\cdot L]$, where $B(t) = e^{zt}$ and $A(T) = e^{xt}$, with $z \geq 0$ and $x \geq 0$. Show that if $z > 0$ and a steady state exists, the production function must take the Cobb–Douglas form.

1.3 Dependence of the saving rate, population growth rate, and depreciation rate on the capital intensity. Assume that the production function satisfies the neoclassical properties.

a. Why would the saving rate, s, generally depend on k? (Provide some intuition; the precise answer will be given in chapter 2.)

b. How does the speed of convergence change if $s(k)$ is an increasing function of k? What if $s(k)$ is a decreasing function of k?

Consider now an AK technology.

c. Why would the saving rate, s, depend on k in this context?

d. How does the growth rate of k change over time depending on whether $s(k)$ is an increasing or decreasing function of k?

e. Suppose that the rate of population growth, n, depends on k. For an AK technology, what would the relation between n and k have to be in order for the model to predict convergence? Can you think of reasons why n would relate to k in this manner? (We analyze the determination of n in chapter 9.)

f. Repeat part e in terms of the depreciation rate, δ. Why might δ depend on k?

1.4 Effects of a higher saving rate. Consider this statement: "Devoting a larger share of national output to investment would help to restore rapid productivity growth and rising living standards." Under what conditions is the statement accurate?

1.5 Factor shares. For a neoclassical production function, show that each factor of production earns its marginal product. Show that if owners of capital save all their income and workers consume all their income, the economy reaches the golden rule of capital accumulation. Explain the results.

1.6 Distortions in the Solow–Swan model (based on Easterly, 1993). Assume that output is produced by the CES production function,

$$Y = \left[\left(a_F K_F^{\eta} + a_I K_I^{\eta}\right)^{\psi/\eta} + a_G K_G^{\psi}\right]^{1/\psi}$$

where Y is output; K_F is formal capital, which is subject to taxation; K_I is informal capital, which evades taxation; K_G is public capital, provided by government and used freely by all producers; $a_F, a_I, a_G > 0$; $\eta < 1$; and $\psi < 1$. Installed formal and informal capital differ in their location and form of ownership and, therefore, in their productivity.

Output can be used on a one-for-one basis for consumption or gross investment in the three types of capital. All three types of capital depreciate at the rate δ. Population is constant, and technological progress is nil.

Formal capital is subject to tax at the rate τ at the moment of its installation. Thus, the price of formal capital (in units of output) is $1 + \tau$. The price of a unit of informal capital is one. Gross investment in public capital is the fixed fraction s_G of tax revenues. Any unused tax receipts are rebated to households in a lump-sum manner. The sum of investment in the two forms of private capital is the fraction s of income net of taxes and transfers. Existing private capital can be converted on a one-to-one basis in either direction between formal and informal capital.

a. Derive the ratio of informal to formal capital used by profit-maximizing producers.

b. In the steady state, the three forms of capital grow at the same rate. What is the ratio of output to formal capital in the steady state?

c. What is the steady-state growth rate of the economy?

d. Numerical simulations show that, for reasonable parameter values, the graph of the growth rate against the tax rate, τ, initially increases rapidly, then reaches a peak, and finally decreases steadily. Explain this nonmonotonic relation between the growth rate and the tax rate.

1.7 A linear production function. Consider the production function $Y = AK + BL$, where A and B are positive constants.

a. Is this production function neoclassical? Which of the neoclassical conditions does it satisfy and which ones does it not?

b. Write output per person as a function of capital per person. What is the marginal product of k? What is the average product of k?

In what follows, we assume that population grows at the constant rate n and that capital depreciates at the constant rate δ.

c. Write down the fundamental equation of the Solow–Swan model.

d. Under what conditions does this model have a steady state with no growth of per capita capital, and under what conditions does the model display endogenous growth?

e. In the case of endogenous growth, how does the growth rate of the capital stock behave over time (that is, does it increase or decrease)? What about the growth rates of output and consumption per capita?

f. If $s = 0.4$, $A = 1$, $B = 2$, $\delta = 0.08$, and $n = 0.02$, what is the long-run growth rate of this economy? What if $B = 5$? Explain the differences.

1.8 Forms of technological progress and steady-state growth. Consider an economy with a CES production function:

$$Y = D(t) \cdot \{[B(t) \cdot K]^{\psi} + [A(t) \cdot L]^{\psi}\}^{1/\psi}$$

where ψ is a constant parameter different from zero. The terms $D(t)$, $B(t)$, and $A(t)$ represent different forms of technological progress. The growth rates of these three terms are constant, and we denote them by x_D, x_B, and x_A, respectively. Assume that population is constant, with $L = 1$, and normalize the initial levels of the three technologies to one, so that $D(0) = B(0) = A(0) = 1$. In this economy, capital accumulates according to the usual equation:

$$\dot{K} = Y - C - \delta K$$

a. Show that, in a steady state (defined as a situation in which all the variables grow at constant, perhaps different, rates), the growth rates of Y, K, and C are the same.

b. Imagine first that $x_B = x_A = 0$ and that $x_D > 0$. Show that the steady state must have $\gamma_K = 0$ (and, therefore, $\gamma_Y = \gamma_C = 0$). (Hint: Show first that $\gamma_Y = x_D + \frac{[K_0 e^{\gamma_K t}]^{\psi}}{1+[K_0 e^{\gamma_K t}]^{\psi}} \cdot \gamma_K$.)

c. Using the results in parts a and b, what is the only growth rate of $D(t)$ that is consistent with a steady state? What, therefore, is the only possible steady-state growth rate of Y?

d. Imagine now that $x_D = x_A = 0$ and that $x_B > 0$. Show that, in the steady state, $\gamma_K = -x_B$ (Hint: Show first that $\gamma_Y = (x_B + \gamma_K) \cdot \frac{[K_t \cdot B_t]^\psi}{1+[K_t \cdot B_t]^\psi}$.)

e. Using the results in parts a and d, show that the only growth rate of B consistent with a steady state is $x_B = 0$.

f. Finally, assume that $x_D = x_B = 0$ and that $x_A > 0$. Show that, in a steady state, the growth rates must satisfy $\gamma_K = \gamma_Y = \gamma_C = x_D$. (Hint: Show first that $\gamma_Y = \frac{K_t^\psi \cdot \gamma_K + A_t^\psi \cdot x_A}{K_t^\psi + A_t^\psi}$.)

g. What would be the steady-state growth rate in part f if population is not constant but, instead, grows at the rate $n > 0$?

2 Growth Models with Consumer Optimization (the Ramsey Model)

One shortcoming of the models that we analyzed in chapter 1 is that the saving rate—and, hence, the ratio of consumption to income—are exogenous and constant. By not allowing consumers to behave optimally, the analysis did not allow us to discuss how incentives affect the behavior of the economy. In particular, we could not think about how the economy reacted to changes in interest rates, tax rates, or other variables. In chapter 1 we showed that allowing for firms to behave optimally did not change any of the basic results of the Solow–Swan model. The main reason was that the overall amount of investment in the economy was still given by the saving of families, and that saving remained exogenous.

To paint a more complete picture of the process of economic growth, we need to allow for the path of consumption and, hence, the saving rate to be determined by optimizing households and firms that interact on competitive markets. We deal here with infinitely lived households that choose consumption and saving to maximize their dynastic utility, subject to an intertemporal budget constraint. This specification of consumer behavior is a key element in the Ramsey growth model, as constructed by Ramsey (1928) and refined by Cass (1965) and Koopmans (1965).

One finding will be that the saving rate is not constant in general but is instead a function of the per capita capital stock, k. Thus we modify the Solow–Swan model in two respects: first, we pin down the average level of the saving rate, and, second, we determine whether the saving rate rises or falls as the economy develops. We also learn how saving rates depend on interest rates and wealth and, in a later chapter, on tax rates and subsidies.

The average level of the saving rate is especially important for the determination of the levels of variables in the steady state. In particular, the optimizing conditions in the Ramsey model preclude the kind of inefficient oversaving that was possible in the Solow–Swan model.

The tendency for saving rates to rise or fall with economic development affects the transitional dynamics, for example, the speed of convergence to the steady state. If the saving rate rises with k, then the convergence speed is slower than that in the Solow–Swan model, and vice versa. We find, however, that even if the saving rate is rising, the convergence property still holds under fairly general conditions in the Ramsey model. That is, an economy still tends to grow faster in per capita terms when it is further from its own steady-state position.

We show that the Solow–Swan model with a constant saving rate is a special case of the Ramsey model; moreover, this case corresponds to reasonable parameter values. Thus, it was worthwhile to begin with the Solow–Swan model as a tractable approximation to the optimizing framework. We also note, however, that the empirical evidence suggests that saving rates typically rise with per capita income during the transition to the steady state. The Ramsey model is consistent with this pattern, and the model allows us to assess the implications of this saving behavior for the transitional dynamics. Moreover, the optimizing framework will be essential in later chapters when we extend the Ramsey model in various respects and consider the possible roles for government policies. Such policies will, in general, affect the incentives to save.

2.1 Households

2.1.1 Setup of the Model

The households provide labor services in exchange for wages, receive interest income on assets, purchase goods for consumption, and save by accumulating assets. The basic model assumes identical households—each has the same preference parameters, faces the same wage rate (because all workers are equally productive), begins with the same assets per person, and has the same rate of population growth. Given these assumptions, the analysis can use the usual representative-agent framework, in which the equilibrium derives from the choices of a single household. We discuss later how the results generalize when various dimensions of household heterogeneity are introduced.

Each household contains one or more adult, working members of the current generation. In making plans, these adults take account of the welfare and resources of their prospective descendants. We model this intergenerational interaction by imagining that the current generation maximizes utility and incorporates a budget constraint over an infinite horizon. That is, although individuals have finite lives, we consider an immortal extended family. This setting is appropriate if altruistic parents provide transfers to their children, who give in turn to their children, and so on. The immortal family corresponds to finite-lived individuals who are connected through a pattern of operative intergenerational transfers based on altruism.[1]

The current adults expect the size of their extended family to grow at the rate n because of the net influences of fertility and mortality. In chapter 9 we study how rational agents choose their fertility by weighing the costs and benefits of rearing children. But, at this point, we continue to simplify by treating n as exogenous and constant. We also neglect migration of persons, another topic explored in chapter 9. If we normalize the number of adults at time 0 to unity, the family size at time t—which corresponds to the adult population—is

$$L(t) = e^{nt}$$

If $C(t)$ is total consumption at time t, then $c(t) \equiv C(t)/L(t)$ is consumption per adult person.

1. See Barro (1974). We abstract from marriage, which generates interactions across family lines. See Bernheim and Bagwell (1988) for a discussion.

Each household wishes to maximize overall utility, U, as given by

$$U = \int_0^{\infty} u[c(t)] \cdot e^{nt} \cdot e^{-\rho t}\, dt \tag{2.1}$$

This formulation assumes that the household's utility at time 0 is a weighted sum of all future flows of utility, $u(c)$. The function $u(c)$—often called the felicity function—relates the flow of utility per person to the quantity of consumption per person, c. We assume that $u(c)$ is increasing in c and concave—$u'(c) > 0$, $u''(c) < 0$.[2] The concavity assumption generates a desire to smooth consumption over time: households prefer a relatively uniform pattern to one in which c is very low in some periods and very high in others. This desire to smooth consumption drives the household's saving behavior because they will tend to borrow when income is relatively low and save when income is relatively high. We also assume that $u(c)$ satisfies Inada conditions: $u'(c) \to \infty$ as $c \to 0$, and $u'(c) \to 0$ as $c \to \infty$.

The multiplication of $u(c)$ in equation (2.1) by family size, $L = e^{nt}$, represents the adding up of utils for all family members alive at time t. The other multiplier, $e^{-\rho t}$, involves the rate of time preference, $\rho > 0$. A positive value of ρ means that utils are valued less the later they are received.[3] We assume $\rho > n$, which implies that U in equation (2.1) is bounded if c is constant over time.

One reason for ρ to be positive is that utils far in the future correspond to consumption of later generations. Suppose that, starting from a point at which the levels of consumption per person in each generation are the same, parents prefer a unit of their own consumption to a unit of their children's consumption. This parental "selfishness" corresponds to $\rho > 0$ in equation (2.1). In a fuller specification, we would also distinguish the rate at which individuals discount their own flow of utility at different points in time (for which $\rho = 0$ might apply) from the rate that applies across generations. Equation (2.1) assumes, only for reasons of tractability, that the discount rate within a person's lifetime is the same as that across generations.

It is also plausible that parents would have diminishing marginal utility with respect to the number of children. We could model this effect by allowing the rate of time preference,

2. The results will be invariant with positive linear transformations of the utility function but not with arbitrary positive, monotonic transformations. Thus, the analysis depends on a limited form of cardinal utility. See Koopmans (1965) for a discussion.

3. Ramsey (1928) preferred to assume $\rho = 0$. He then interpreted the optimizing agent as a social planner, rather than a competitive household, who chose consumption and saving for today's generation as well as for future generations. The discounting of utility for future generations ($\rho > 0$) was, according to Ramsey, "ethically indefensible." We work out an example with $\rho = 0$ in the mathematics chapter.

ρ, to increase with the population growth rate, n.[4] Because we treat n as exogenous, this dependence of ρ on n would not materially change the analysis in this chapter. We shall, however, consider this effect in chapter 9, which allows for an endogenous determination of population growth.

Households hold assets in the form of ownership claims on capital (to be introduced later) or as loans. Negative loans represent debts. We continue to assume a closed economy, so that no assets can be traded internationally. Households can lend to and borrow from other households, but the representative household will end up holding zero net loans in equilibrium. Because the two forms of assets, capital and loans, are assumed to be perfect substitutes as stores of value, they must pay the same real rate of return, $r(t)$. We denote the household's net assets per person by $a(t)$, where $a(t)$ is measured in real terms, that is, in units of consumables.

Households are competitive in that each takes as given the interest rate, $r(t)$, and the wage rate, $w(t)$, paid per unit of labor services. We assume that each adult supplies inelastically one unit of labor services per unit of time. (Chapter 9 considers a labor/leisure choice.) In equilibrium, the labor market clears, and the household obtains the desired quantity of employment. That is, the model abstracts from "involuntary unemployment."

Since each person works one unit of labor services per unit of time, the wage income per adult person equals $w(t)$. The total income received by the aggregate of households is, therefore, the sum of labor income, $w(t) \cdot L(t)$, and asset income, $r(t) \cdot (\text{Assets})$. Households use the income that they do not consume to accumulate more assets:

$$\frac{d(\text{Assets})}{dt} = r \cdot (\text{Assets}) + wL - C \tag{2.2}$$

where we omit time subscripts whenever no ambiguity results. Since a is per capita assets, we have

$$\dot{a} = \left(\frac{1}{L}\right) \cdot \left[\frac{d(\text{Assets})}{dt}\right] - na$$

Therefore, if we divide equation (2.2) by L, we get the budget constraint in per capita terms:

$$\dot{a} = w + ra - c - na \tag{2.3}$$

4. One case common in the growth literature assumes that ρ rises one to one with n; that is, $\rho = \rho^* + n$, where ρ^* is the positive rate of time preference that applies under zero population growth. In this case, utility at time t enters into equation (2.1) as $u(c)e^{-\rho^* t}$, which depends on per capita utility, but not on the size of the family at time t. This specification is used, for example, by Sidrauski (1967) and Blanchard and Fischer (1989, chapter 2).

If each household can borrow an unlimited amount at the going interest rate, $r(t)$, it has an incentive to pursue a form of chain letter or Ponzi game. The household can borrow to finance current consumption and then use future borrowings to roll over the principal and pay all the interest. In this case, the household's debt grows forever at the rate of interest, $r(t)$. Since no principal ever gets repaid, today's added consumption is effectively free. Thus a household that can borrow in this manner would be able to finance an arbitrarily high level of consumption in perpetuity.

To rule out chain-letter possibilities, we assume that the credit market imposes a constraint on the amount of borrowing. The appropriate restriction turns out to be that the present value of assets must be asymptotically nonnegative, that is,

$$\lim_{t\to\infty}\left\{a(t)\cdot\exp\left[-\int_0^t [r(v)-n]\,dv\right]\right\}\geq 0 \tag{2.4}$$

This constraint means that, in the long run, a household's debt per person (negative values of $a[t]$) cannot grow as fast as $r(t)-n$, so that the level of debt cannot grow as fast as $r(t)$. This restriction rules out the type of chain-letter finance that we have described. We show later how the credit-market constraint expressed in equation (2.4) emerges naturally from the market equilibrium.

The household's optimization problem is to maximize U in equation (2.1), subject to the budget constraint in equation (2.3), the stock of initial assets, $a(0)$, and the limitation on borrowing in equation (2.3). The inequality restrictions, $c(t)\geq 0$, also apply. However, as $c(t)$ approaches 0, the Inada condition implies that the marginal utility of consumption becomes infinite. The inequality restrictions will therefore never bind, and we can safely ignore them.

2.1.2 First-Order Conditions

The mathematical methods for this type of dynamic optimization problem are discussed in the appendix on mathematics at the end of the book. We use these results here without further derivation. Begin with the present-value Hamiltonian,

$$J = u[c(t)]\cdot e^{-(\rho-n)t}+\nu(t)\cdot\{w(t)+[r(t)-n]\cdot a(t)-c(t)\} \tag{2.5}$$

where the expression in braces equals $\dot{a}$ from equation (2.3). The variable $\nu(t)$ is the present-value shadow price of income. It represents the value of an increment of income received at time t in units of utils at time 0.[5] Notice that this shadow price depends on time because there

5. We could deal alternatively with the shadow price $\nu e^{(\rho-n)t}$. This shadow price measures the value of an increment of income at time t in units of utils at time t. (See the discussion in the appendix on mathematics at the end of the book.)

is one of them for each "constraint," and the household faces a continuum of constraints, one for each instant. The first-order conditions for a maximum of U are

$$\frac{\partial J}{\partial c} = 0 \Longrightarrow \nu = u'(c)e^{-(\rho-n)t} \tag{2.6}$$

$$\dot{\nu} = -\partial J/\partial a \Longrightarrow \dot{\nu} = -(r-n)\cdot\nu \tag{2.7}$$

The transversality condition is

$$\lim_{t\to\infty}[\nu(t)\cdot a(t)] = 0 \tag{2.8}$$

The Euler Equation If we differentiate equation (2.6) with respect to time and substitute for ν from this equation and for $\dot{\nu}$ from equation (2.7), we get the basic condition for choosing consumption over time:

$$r = \rho - \left(\frac{du'/dt}{u'}\right) = \rho - \left[\frac{u''(c)\cdot c}{u'(c)}\right]\cdot(\dot{c}/c) \tag{2.9}$$

This equation says that households choose consumption so as to equate the rate of return, r, to the rate of time preference, ρ, plus the rate of decrease of the marginal utility of consumption, u', due to growing per capita consumption, c.

The interest rate, r, on the left-hand side of equation (2.9) is the rate of return to saving. The far right-hand side of the equation can be viewed as the rate of return to consumption. Agents prefer to consume today rather than tomorrow for two reasons. First, because households discount future utility at rate ρ, this rate is part of the rate of return to consumption today. Second, if $\dot{c}/c > 0$, c is low today relative to tomorrow. Since agents like to smooth consumption over time—because $u''(c) < 0$—they would like to even out the flow by bringing some future consumption forward to the present. The second term on the far right picks up this effect. If agents are optimizing, equation (2.9) says that they have equated the two rates of return and are therefore indifferent at the margin between consuming and saving.

Another way to view equation (2.9) is that households would select a flat consumption profile, with $\dot{c}/c = 0$, if $r = \rho$. Households would be willing to depart from this flat pattern and sacrifice some consumption today for more consumption tomorrow—that is, tolerate $\dot{c}/c > 0$—only if they are compensated by an interest rate, r, that is sufficiently above ρ. The term $[\frac{-u''(c)\cdot c}{u'(c)}]\cdot(\dot{c}/c)$ on the right-hand side of equation (2.9) gives the required amount of compensation. Note that the term in brackets is the magnitude of the elasticity of $u'(c)$ with respect to c. This elasticity, a measure of the concavity of $u(c)$, determines the amount by which r must exceed ρ. If the elasticity is larger in magnitude, the required premium of r over ρ is greater for a given value of $\dot{c}/c$.

The magnitude of the elasticity of marginal utility, $\{[-u''(c) \cdot c]/[u'(c)]\}$, is sometimes called the reciprocal of the elasticity of intertemporal substitution.[6] Equation (2.9) shows that to find a steady state in which r and $\dot{c}/c$ are constant, this elasticity must be constant asymptotically. We therefore follow the common practice of assuming the functional form

$$u(c) = \frac{c^{(1-\theta)} - 1}{(1-\theta)} \tag{2.10}$$

where $\theta > 0$, so that the elasticity of marginal utility equals the constant $-\theta$.[7] The elasticity of substitution for this utility function is the constant $\sigma = 1/\theta$. Hence, this form is called the *constant intertemporal elasticity of substitution* (CIES) utility function. The higher is θ, the more rapid is the proportionate decline in $u'(c)$ in response to increases in c and, hence, the less willing households are to accept deviations from a uniform pattern of c over time. As θ approaches 0, the utility function approaches a linear form in c; the linearity means that households are indifferent to the timing of consumption if $r = \rho$ applies.

The form of $u(c)$ in equation (2.10) implies that the optimality condition from equation (2.9) simplifies to

$$\dot{c}/c = (1/\theta) \cdot (r - \rho) \tag{2.11}$$

Therefore, the relation between r and ρ determines whether households choose a pattern of per capita consumption that rises over time, stays constant, or falls over time. A lower willingness to substitute intertemporally (a higher value of θ) implies a smaller responsiveness of $\dot{c}/c$ to the gap between r and ρ.

The Transversality Condition The transversality condition in equation (2.8) says that the value of the household's per capita assets—the quantity $a(t)$ times the shadow price

6. The elasticity of intertemporal substitution between consumption at times t_1 and t_2 is given by the reciprocal of the proportionate change in the magnitude of the slope of an indifference curve in response to a proportionate change in the ratio $c(t_1)/c(t_2)$. If we denote this elasticity by σ, we get

$$\sigma = \left[\frac{c(t_1)/c(t_2)}{-u'[c(t_1)]/u'[c(t_2)]} \cdot \frac{d\{u'[c(t_1)]/u'[c(t_2)]\}}{d[c(t_1)/c(t_2)]}\right]^{-1}$$

where $-u'[c(t_1)]/u'[c(t_2)]$ is the magnitude of the slope of the indifference curve. If we let t_2 approach t_1, we get the instantaneous elasticity,

$$\sigma = -u'(c)/[c \cdot u''(c)]$$

which is the inverse of the magnitude of the elasticity of marginal utility.

7. The inclusion of the -1 in the formula is convenient because it implies that $u(c)$ approaches $\log(c)$ as $\theta \to 1$. (This result can be proven using l'Hôpital's rule.) The term $-1/(1-\theta)$ can, however, be omitted without affecting the subsequent results, because the household's choices are invariant with respect to linear transformations of the utility function (see footnote 2).

$\nu(t)$—must approach 0 as time approaches infinity. If we think of infinity loosely as the end of the planning horizon, the intuition is that optimizing agents do not want to have any valuable assets left over at the end.[8] Utility would increase if the assets, which are effectively being wasted, were used instead to raise consumption at some dates in finite time.

The shadow price ν evolves over time in accordance with equation (2.7). Integration of this equation with respect to time yields

$$\nu(t) = \nu(0) \cdot \exp\left\{-\int_0^t [r(v) - n]\, dv\right\}$$

The term $\nu(0)$ equals $u'[c(0)]$, which is positive because $c(0)$ is finite (if U is finite), and $u'(c)$ is assumed to be positive as long as c is finite.

If we substitute the result for $\nu(t)$ into equation (2.8), the transversality condition becomes

$$\lim_{t\to\infty} \left\{a(t) \cdot \exp\left[-\int_0^t [r(v) - n]\, dv\right]\right\} = 0 \tag{2.12}$$

This equation implies that the quantity of assets per person, a, does not grow asymptotically at a rate as high as $r - n$ or, equivalently, that the level of assets does not grow at a rate as high as r. It would be suboptimal for households to accumulate positive assets forever at the rate r or higher, because utility would increase if these assets were instead consumed in finite time.

In the case of borrowing, where $a(t)$ is negative, infinite-lived households would like to violate equation (2.12) by borrowing and never making payments for principal or interest. However, equation (2.4) rules out this chain-letter finance, that is, schemes in which a household's debt grows forever at the rate r or higher. In order to borrow on this perpetual basis, households would have to find willing lenders; that is, other households that were willing to hold positive assets that grew at the rate r or higher. But we already know from the transversality condition that these other households will be unwilling to absorb assets asymptotically at such a high rate. Therefore, in equilibrium, each household will be unable to borrow in a chain-letter fashion. In other words, the inequality restriction shown in equation (2.4) is not arbitrary and would, in fact, be imposed in equilibrium by the credit market. Faced by this constraint, the best thing that optimizing households can do is to satisfy the condition shown in equation (2.12). That is, this equality holds whether $a(t)$ is positive or negative.

8. The interpretation of the transversality condition in the infinite-horizon problem as the limit of the corresponding condition for a finite-horizon problem is not always correct. See the appendix on mathematics at the end of the book.

The Consumption Function The term $\exp[-\int_0^t r(v)\,dv]$, which appears in equation (2.12), is a present-value factor that converts a unit of income at time t to an equivalent unit of income at time 0. If $r(v)$ equaled the constant r, the present-value factor would simplify to e^{-rt}. More generally we can think of an average interest rate between times 0 and t, defined by

$$\bar{r}(t) = (1/t) \cdot \int_0^t r(v)\,dv \tag{2.13}$$

The present-value factor equals $e^{-\bar{r}(t)\cdot t}$.

Equation (2.11) determines the growth rate of c. To determine the level of c—that is, the consumption function—we have to use the flow budget constraint, equation (2.3), to derive the household's intertemporal budget constraint. We can solve equation (2.3) as a first-order linear differential equation in a to get an intertemporal budget constraint that holds for any time $T \geq 0$:[9]

$$a(T) \cdot e^{-[\bar{r}(T)-n]T} + \int_0^T c(t)e^{-[\bar{r}(t)-n]t}\,dt = a(0) + \int_0^T w(t)e^{-[\bar{r}(t)-n]t}\,dt$$

where we used the definition of $\bar{r}(t)$ from equation (2.13). This intertemporal budget constraint says that the present discounted value of all income between 0 and T plus the initial available wealth have to equal the present discounted value of all future consumption plus the present value of the assets left at T. If we take the limit as $T \to \infty$, the term on the far left vanishes (from the transversality condition in equation [2.12]), and the intertemporal budget constraint becomes

$$\int_0^\infty c(t)e^{-[\bar{r}(t)-n]t}\,dt = a(0) + \int_0^\infty w(t)e^{-[\bar{r}(t)-n]t}\,dt = a(0) + \tilde{w}(0) \tag{2.14}$$

Hence, the present value of consumption equals lifetime wealth, defined as the sum of initial assets, $a(0)$, and the present value of wage income, denoted by $\tilde{w}(0)$.

If we integrate equation (2.11) between times 0 and t and use the definition of $\bar{r}(t)$ from equation (2.13), we find that consumption is given by

$$c(t) = c(0) \cdot e^{(1/\theta)\cdot[\bar{r}(t)-\rho]t}$$

9. The methods for solving first-order linear differential equations with variable coefficients are discussed in the appendix on mathematics at the end of the book.

Substitution of this result for $c(t)$ into the intertemporal budget constraint in equation (2.14) leads to the consumption function at time 0:

$$c(0) = \mu(0) \cdot [a(0) + \tilde{w}(0)] \tag{2.15}$$

where $\mu(0)$, the propensity to consume out of wealth, is determined from

$$[1/\mu(0)] = \int_0^\infty e^{[\bar{r}(t)\cdot(1-\theta)/\theta-\rho/\theta+n]t}\, dt \tag{2.16}$$

An increase in average interest rates, $\bar{r}(t)$, for given wealth, has two effects on the marginal propensity to consume in equation (2.16). First, higher interest rates increase the cost of current consumption relative to future consumption, an intertemporal-substitution effect that motivates households to shift consumption from the present to the future. Second, higher interest rates have an income effect that tends to raise consumption at all dates. The net effect of an increase in $\bar{r}(t)$ on $\mu(0)$ depends on which of the two forces dominates.

If $\theta < 1$, $\mu(0)$ declines with $\bar{r}(t)$ because the substitution effect dominates. The intuition is that, when θ is low, households care relatively little about consumption smoothing, and the intertemporal-substitution effect is large. Conversely, if $\theta > 1$, $\mu(0)$ rises with $\bar{r}(t)$ because the substitution effect is relatively weak. Finally, if $\theta = 1$ (log utility), the two effects exactly cancel, and $\mu(0)$ simplifies to $\rho - n$, which is independent of $\bar{r}(t)$. Recall that we assumed $\rho - n > 0$.

The effects of $\bar{r}(t)$ on $\mu(0)$ carry over to effects on $c(0)$ if we hold constant the wealth term, $a(0) + \tilde{w}(0)$. In fact, however, $\tilde{w}(0)$ falls with $\bar{r}(t)$ for a given path of $w(t)$. This third effect reinforces the substitution effect that we mentioned before.

2.2 Firms

Firms produce goods, pay wages for labor input, and make rental payments for capital input. Each firm has access to the production technology,

$$Y(t) = F[K(t), L(t), T(t)]$$

where Y is the flow of output, K is capital input (in units of commodities), L is labor input (in person-hours per year), and $T(t)$ is the level of the technology, which is assumed to grow at the constant rate $x \geq 0$. Hence, $T(t) = e^{xt}$, where we normalize the initial level of technology, $T(0)$, to 1. The function $F(\cdot)$ satisfies the neoclassical properties discussed in chapter 1. In particular, Y exhibits constant returns to scale in K and L, and each input exhibits positive and diminishing marginal product.

We showed in chapter 1 that a steady state coexists with technological progress at a constant rate only if this progress takes the labor-augmenting form

$$Y(t) = F[K(t), L(t) \cdot T(t)]$$

If we again define "effective labor" as the product of raw labor and the level of technology, $\hat{L} \equiv L \cdot T(t)$, the production function can be written as

$$Y = F(K, \hat{L}) \tag{2.17}$$

We shall find it convenient to work with variables that are constant in the steady state. In chapter 1, we showed that the steady state of the model with exogenous technical progress was such that the per capita variables grew at the rate of technological progress, x. This property will still hold in the present model. Hence, we will deal again with quantities per unit of effective labor:

$$\hat{y} \equiv Y/\hat{L} \text{ and } \hat{k} \equiv K/\hat{L}$$

The production function can then be rewritten in intensive form, as in equation (1.38),

$$\hat{y} = f(\hat{k}) \tag{2.18}$$

where $f(0) = 0$. It can be readily verified that the marginal products of the factors are given by[10]

$$\partial Y/\partial K = f'(\hat{k})$$

$$\partial Y/\partial L = [f(\hat{k}) - \hat{k} \cdot f'(\hat{k})] \cdot e^{xt} \tag{2.19}$$

The Inada conditions, discussed in chapter 1, imply $f'(\hat{k}) \to \infty$ as $\hat{k} \to 0$ and $f'(\hat{k}) \to 0$ as $\hat{k} \to \infty$.

We think of firms as renting the services of capital from the households that own the capital. (None of the results would change if the firms owned the capital, and the households owned shares of stock in the firms.) If we let $R(t)$ be the rental rate of a unit of capital, a firm's total cost for capital is RK, which is proportional to K. We assume that capital services can be increased or decreased without incurring any additional expenses, such as costs for installing machines or making other changes. We consider these kinds of adjustment costs in chapter 3.

We assume, as in chapter 1, a one-sector production model in which one unit of output can be used to generate one unit of household consumption, C, or one unit of additional

10. We can write $Y = \hat{L} \cdot f(\hat{k})$. Differentiation of Y with respect to K, holding fixed L and t, leads to $\partial Y/\partial K = f'(\hat{k})$. Differentiation of Y with respect to L, holding fixed K and t, leads to $\partial Y/\partial L = [f(\hat{k}) - \hat{k} \cdot f'(\hat{k})]e^{xt}$.

capital, K. Therefore, as long as the economy is not at a corner solution in which all current output goes into consumption or new capital, the price of K in terms of C will be fixed at unity. Because C will be nonzero in equilibrium, we have to be concerned only with the possibility that none of the output goes into new capital; in other words, that gross investment is 0. Even in this situation, the price of K in terms of C would remain at unity if capital were reversible in the sense that the existing stocks could be consumed on a one-for-one basis. With reversible capital, the economy's gross investment can be negative, and the price of K in units of C stays at unity. Although this situation may apply to farm animals, economists usually assume that investment is irreversible. In this case, the price of K in units of C is one only if the constraint of nonnegative aggregate gross investment is nonbinding in equilibrium. We maintain this assumption in the following analysis, and we deal with irreversible investment in appendix 2B (section 2.9).

Since capital stocks depreciate at the constant rate $\delta \geq 0$, the net rate of return to a household that owns a unit of capital is $R - \delta$.[11] Recall that households can also receive the interest rate r on funds lent to other households. Since capital and loans are perfect substitutes as stores of value, we must have $r = R - \delta$ or, equivalently, $R = r + \delta$.

The representative firm's flow of net receipts or profit at any point in time is given by

$$\pi = F(K, \hat{L}) - (r + \delta) \cdot K - wL \tag{2.20}$$

As in chapter 1, the problem of maximizing the present value of profit reduces here to a problem of maximizing profit in each period without regard to the outcomes in other periods. Profit can be written as

$$\pi = \hat{L} \cdot [f(\hat{k}) - (r + \delta) \cdot \hat{k} - we^{-xt}] \tag{2.21}$$

A competitive firm, which takes r and w as given, maximizes profit for given $\hat{L}$ by setting

$$f'(\hat{k}) = r + \delta \tag{2.22}$$

Also as before, in a full-market equilibrium, w equals the marginal product of labor corresponding to the value of $\hat{k}$ that satisfies equation (2.22):

$$[f(\hat{k}) - \hat{k} \cdot f'(\hat{k})]e^{xt} = w \tag{2.23}$$

This condition ensures that profit equals zero for any value of $\hat{L}$.

11. More generally, if the price of capital can change over time, the real rate of return for owners of capital equals $R/\phi - \delta + \dot{\phi}/\phi$, where ϕ is the price of capital in units of consumables. In the present case, where $\phi = 1$, the capital-gain term, $\dot{\phi}/\phi$, vanishes, and the rate of return simplifies to $R - \delta$.

2.3 Equilibrium

We began with the behavior of competitive households that faced a given interest rate, r, and wage rate, w. We then introduced competitive firms that also faced given values of r and w. We can now combine the behavior of households and firms to analyze the structure of a competitive market equilibrium.

Since the economy is closed, all debts within the economy must cancel. Hence, the assets per adult person, a, equal the capital per worker, k. The equality between k and a follows because all of the capital stock must be owned by someone in the economy; in particular, in this closed-economy model, all of the domestic capital stock must be owned by the domestic residents. If the economy were open to international capital markets, the gap between k and a would correspond to the home country's net debt to foreigners. Chapter 3 considers an open economy, in which the net foreign debt can be nonzero.

The household's flow budget constraint in equation (2.3) determines $\dot{a}$. Use $a = k$, $\hat{k} = ke^{-xt}$, and the conditions for r and w in equations (2.22) and (2.23) to get

$$\dot{\hat{k}} = f(\hat{k}) - \hat{c} - (x + n + \delta) \cdot \hat{k} \tag{2.24}$$

where $\hat{c} \equiv C/\hat{L} = ce^{-xt}$, and $\hat{k}(0)$ is given. Equation (2.24) is the resource constraint for the overall economy: the change in the capital stock equals output less consumption and depreciation, and the change in $\hat{k} \equiv K/\hat{L}$ also takes account of the growth in $\hat{L}$ at the rate $x + n$.

The differential equation (2.24) is the key relation that determines the evolution of $\hat{k}$ and, hence, $\hat{y} = f(\hat{k})$ over time. The missing element, however, is the determination of $\hat{c}$. If we knew the relation of $\hat{c}$ to $\hat{k}$ (or $\hat{y}$), or if we had another differential equation that determined the evolution of $\hat{c}$, we could study the full dynamics of the economy.

In the Solow–Swan model of chapter 1, the missing relation was provided by the assumption of a constant saving rate. This assumption implied the linear consumption function, $\hat{c} = (1 - s) \cdot f(\hat{k})$. In the present setting, the behavior of the saving rate is not so simple, but we do know from household optimization that c grows in accordance with equation (2.11). If we use the conditions $r = f'(\hat{k}) - \delta$ and $\hat{c} = ce^{-xt}$, we get

$$\dot{\hat{c}}/\hat{c} = \frac{\dot{c}}{c} - x = \frac{1}{\theta} \cdot [f'(\hat{k}) - \delta - \rho - \theta x] \tag{2.25}$$

This equation, together with equation (2.24), forms a system of two differential equations in $\hat{c}$ and $\hat{k}$. This system, together with the initial condition, $\hat{k}(0)$, and the transversality condition, determines the time paths of $\hat{c}$ and $\hat{k}$.

We can write the transversality condition in terms of $\hat{k}$ by substituting $a = k$ and $\hat{k} = ke^{-xt}$ into equation (2.12) to get

$$\lim_{t\to\infty} \left\{ \hat{k} \cdot \exp\left(-\int_0^t [f'(\hat{k}) - \delta - x - n]\, dv \right) \right\} = 0 \tag{2.26}$$

We can interpret this result if we jump ahead to use the result that $\hat{k}$ tends asymptotically to a constant steady-state value, $\hat{k}^*$, just as in the Solow–Swan model. The transversality condition in equation (2.26) therefore requires $f'(\hat{k}^*) - \delta$, the steady-state rate of return, to exceed $x + n$, the steady-state growth rate of K.

2.4 Alternative Environments

The analysis applies thus far to a decentralized economy with competitive households and firms. We can see from the setup of the model, however, that the same equations—and, hence, the same results—would emerge under some alternative environments. First, households could perform the functions of firms by employing adult family members as workers in accordance with the production process, $f(\hat{k})$.[12] The resource constraint in equation (2.24) follows directly (total output must be allocated to consumption or gross investment, which equals net investment plus depreciation). If the households maximize the utility function in equations (2.1) and (2.10), subject to equation (2.24), then equations (2.25) and (2.26) still represent the first-order conditions. Thus, the separation of functions between households and firms is not central to the analysis.

We could also pretend that the economy was run by a benevolent *social planner,* who dictates the choices of consumption over time and who seeks to maximize the utility of the representative family. The device of the benevolent social planner will be useful in many circumstances for finding the economy's first-best outcomes. The planner is assumed to have the same form of preferences as those assumed before—in particular, the same rate of time preference, ρ, and the same utility function, $u(c)$. The planner is also constrained by the aggregate resource constraint in equation (2.24). The solution for the planner will therefore be the same as that for the decentralized economy.[13] Since a benevolent

12. This setup was considered in chapter 1.

13. The planner's problem is to choose the path of c to maximize U in equation (2.1), subject to the economy's budget constraint in equation (2.24), the initial value $\hat{k}(0)$, and the inequalities $c \geq 0$ and $\hat{k} \geq 0$. The Hamiltonian for this problem is

$$J = u(c)e^{-\rho t} + \nu \cdot [f(\hat{k}) - ce^{-xt} - (x + n + \delta) \cdot \hat{k}]$$

The usual first-order conditions lead to equation (2.25), and the transversality condition leads to equation (2.26).

social planner with dictatorial powers will attain a Pareto optimum, the results for the decentralized economy—which coincide with those of the planner—must also be Pareto optimal.

2.5 The Steady State

We now consider whether the equilibrium conditions, equations (2.24), (2.25), and (2.26), are consistent with a steady state, that is, a situation in which the various quantities grow at constant (possibly zero) rates. We show first that the steady-state growth rates of $\hat{k}$ and $\hat{c}$ must be zero, just as in the Solow–Swan model of chapter 1.

Let $(\gamma_{\hat{k}})^*$ be the steady-state growth rate of $\hat{k}$ and $(\gamma_{\hat{c}})^*$ the steady-state growth rate of $\hat{c}$. In the steady state, equation (2.25) implies

$$\hat{c} = f(\hat{k}) - (x + n + \delta) \cdot \hat{k} - \hat{k} \cdot (\gamma_{\hat{k}})^* \tag{2.27}$$

If we differentiate this condition with respect to time, we find that

$$\dot{\hat{c}} = \dot{\hat{k}} \cdot \{f'(\hat{k}) - [x + n + \delta + (\gamma_{\hat{k}})^*]\} \tag{2.28}$$

must hold in the steady state. The expression in the large braces is positive from the transversality condition shown in equation (2.26). Therefore, $(\gamma_{\hat{k}})^*$ and $(\gamma_{\hat{c}})^*$ must have the same sign.

If $(\gamma_{\hat{k}})^* > 0$, $\hat{k} \to \infty$ and $f'(\hat{k}) \to 0$. Equation (2.25) then implies $(\gamma_{\hat{c}}) < 0$, an outcome that contradicts the result that $(\gamma_{\hat{k}})^*$ and $(\gamma_{\hat{c}})^*$ are of the same sign. If $(\gamma_{\hat{k}})^* < 0$, $\hat{k} \to 0$ and $f'(\hat{k}) \to \infty$. Equation (2.25) then implies $(\gamma_{\hat{c}})^* > 0$, an outcome that again contradicts the result that $(\gamma_{\hat{k}})^*$ and $(\gamma_{\hat{c}})^*$ are of the same sign. Therefore, the only remaining possibility is $(\gamma_{\hat{k}})^* = (\gamma_{\hat{c}})^* = 0$. The result $(\gamma_{\hat{k}})^* = 0$ implies $(\gamma_{\hat{y}})^* = 0$. Thus the variables per unit of effective labor, $\hat{k}$, $\hat{c}$, and $\hat{y}$, are constant in the steady state. This behavior implies that the per capita variables, k, c, and y, grow in the steady state at the rate x, and the level variables, K, C, and Y, grow in the steady state at the rate $x + n$. These results on steady-state growth rates are the same as those in the Solow–Swan model, in which the saving rate was exogenous and constant.

The steady-state values for $\hat{c}$ and $\hat{k}$ are determined by setting the expressions in equations (2.24) and (2.25) to zero. The solid curve in figure 2.1, which corresponds to $\hat{c} = f(\hat{k}) - (x + n + \delta) \cdot \hat{k}$, shows pairs of $(\hat{k}, \hat{c})$ that satisfy $\dot{\hat{k}} = 0$ in equation (2.24). Note that the peak in the curve occurs when $f'(\hat{k}) = \delta + x + n$, so that the interest rate, $f'(\hat{k}) - \delta$, equals the steady-state growth rate of output, $x + n$. This equality between the interest rate

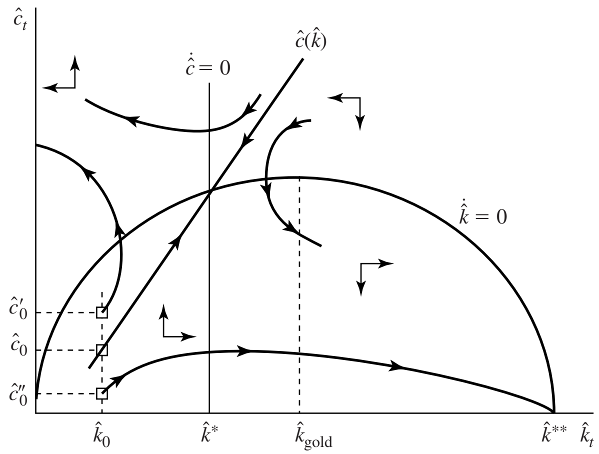

Figure 2.1
The phase diagram of the Ramsey model. The figure shows the transitional dynamics of the Ramsey model. The $\dot{\hat{c}}/\hat{c} = 0$ and $\dot{\hat{k}} = 0$ loci divide the space into four regions, and the arrows show the directions of motion in each region. The model exhibits saddle-path stability. The stable arm is an upward-sloping curve that goes through the origin and the steady state. Starting from a low level of $\hat{k}$, the optimal initial $\hat{c}$ is low. Along the transition, $\hat{c}$ and $\hat{k}$ increase toward their steady-state values.

and the growth rate corresponds to the golden-rule level of $\hat{k}$ (as described in chapter 1),[14] because it leads to a maximum of $\hat{c}$ in the steady state. We denote by $\hat{k}_{gold}$ the value of $\hat{k}$ that corresponds to the golden rule.

Equation (2.25) and the condition $\dot{\hat{c}} = 0$ imply

$$f'(\hat{k}^*) = \delta + \rho + \theta x \tag{2.29}$$

This equation says that the steady-state interest rate, $f'(\hat{k}) - \delta$, equals the effective discount rate, $\rho + \theta x$.[15] The vertical line at $\hat{k}^*$ in figure 2.1 corresponds to this condition; note that $\dot{\hat{c}}/\hat{c} = 0$ holds at this value of $\hat{k}$ independently of the value of $\hat{c}$.[16] The key to the determination of $\hat{k}^*$ in equation (2.29) is the diminishing returns to capital, which make $f'(\hat{k}^*)$ a

14. In chapter 1 we defined the golden-rule level of k as the capital stock per person that maximizes steady-state consumption per capita. It was shown that this level of capital was such that $f'(k_{gold}) = \delta + n$; see equation (1.22). When exogenous technological progress exists, the golden-rule level of $\hat{k}$ is defined as the level that maximizes steady-state consumption per effective unit of labor, $\hat{c} = f(\hat{k}) - (x + n + \delta) \cdot \hat{k}$. Notice that the maximum is achieved when $f'(\hat{k}_{gold}) = (x + n + \delta)$.

15. The θx part of the effective discount rate picks up the effect from diminishing marginal utility of consumption due to growth of c at the rate x. See equation (2.9).

16. Equation (2.25) indicates that $\dot{\hat{c}}/\hat{c} = 0$ is also satisfied if $\hat{c} = 0$, that is, along the horizontal axis in figure 2.1.

monotonically decreasing function of $\hat{k}^*$. Moreover, the Inada conditions—$f'(0) = \infty$ and $f'(\infty) = 0$—ensure that equation (2.29) holds at a unique positive value of $\hat{k}^*$.

Figure 2.1 shows the determination of the steady-state values, $(\hat{k}^*, \hat{c}^*)$, at the intersection of the vertical line with the solid curve. In particular, with $\hat{k}^*$ determined from equation (2.29), the value for $\hat{c}^*$ follows from setting the expression in equation (2.24) to 0 as

$$\hat{c}^* = f(\hat{k}^*) - (x + n + \delta) \cdot \hat{k}^* \tag{2.30}$$

Note that $\hat{y}^* = f(\hat{k}^*)$ is the steady-state value of $\hat{y}$.

Consider the transversality condition in equation (2.26). Since $\hat{k}$ is constant in the steady state, this condition holds if the steady-state rate of return, $r^* = f'(\hat{k}^*) - \delta$, exceeds the steady-state growth rate, $x + n$. Equation (2.29) implies that this condition can be written as

$$\rho > n + (1 - \theta)x \tag{2.31}$$

If ρ is not high enough to satisfy equation (2.31), the household's optimization problem is not well posed because infinite utility would be attained if c grew at the rate x.[17] We assume henceforth that the parameters satisfy equation (2.31).

In figure 2.1, the steady-state value, $\hat{k}^*$, was drawn to the left of $\hat{k}_{gold}$. This relation always holds if the transversality condition, equation (2.31), is satisfied. The steady-state value is determined from $f'(\hat{k}^*) = \delta + \rho + \theta x$,[18] whereas the golden-rule value comes from $f'(\hat{k}_{gold}) = \delta + x + n$. The inequality in equation (2.31) implies $\rho + \theta x > x + n$ and, hence, $f'(\hat{k}^*) > f'(\hat{k}_{gold})$. The result $\hat{k}^* < \hat{k}_{gold}$ follows from $f''(\hat{k}) < 0$.

The implication is that inefficient oversaving cannot occur in the optimizing framework, although it could arise in the Solow–Swan model with an arbitrary, constant saving rate. If the infinitely lived household were oversaving, it would realize that it was not optimizing—because it was not satisfying the transversality condition—and would therefore shift to a path that entailed less saving. Note that the optimizing household does not save enough to attain the golden-rule value, $\hat{k}_{gold}$. The reason is that the impatience reflected in the effective discount rate, $\rho + \theta x$, makes it not worthwhile to sacrifice more of current consumption to reach the maximum of $\hat{c}$ (the golden-rule value, $\hat{c}_{gold}$) in the steady state.

The steady-state growth rates do not depend on parameters that describe the production function, $f(\cdot)$, or on the preference parameters, ρ and θ, that characterize households' attitudes about consumption and saving. These parameters do have long-run effects on levels of variables.

17. The appendix on mathematics at the end of the book considers some cases in which infinite utility can be handled.

18. This condition is sometimes called the *modified golden rule*.

In figure 2.1, an increased willingness to save—represented by a reduction in ρ or θ—shifts the $\dot{\hat{c}}/\hat{c} = 0$ schedule to the right and leaves the $\dot{\hat{k}} = 0$ schedule unchanged. These shifts lead accordingly to higher values of $\hat{c}^*$ and $\hat{k}^*$ and, hence, to a higher value of $\hat{y}^*$. Similarly, a proportional upward shift of the production function or a reduction of the depreciation rate, δ, moves the $\dot{\hat{k}} = 0$ curve up and the $\dot{\hat{c}}/\hat{c} = 0$ curve to the right. These shifts generate increases in $\hat{c}^*$, $\hat{k}^*$, and $\hat{y}^*$. An increase in x raises the effective time-preference term, $\rho + \theta x$, in equation (2.29) and also lowers the value of $\hat{c}^*$ that corresponds to a given $\hat{k}^*$ in equation (2.30). In figure 2.1, these changes shift the $\dot{\hat{k}} = 0$ schedule downward and the $\dot{\hat{c}}/\hat{c} = 0$ schedule leftward and thereby reduce $\hat{c}^*$, $\hat{k}^*$, and $\hat{y}^*$. (Although $\hat{c}$ falls, utility rises because the increase in x raises the growth rate of c relative to that of $\hat{c}$.) Finally, the effect of n on $\hat{k}^*$ and $\hat{y}^*$ is nil if we hold fixed ρ. Equation (2.30) implies that $\hat{c}^*$ declines. If a higher n leads to a higher rate of time preference (for reasons discussed before), then an increase in n would reduce $\hat{k}^*$ and $\hat{y}^*$.

2.6 Transitional Dynamics

2.6.1 The Phase Diagram

The Ramsey model, like the Solow–Swan model, is most interesting for its predictions about the behavior of growth rates and other variables along the transition path from an initial factor ratio, $\hat{k}(0)$, to the steady-state ratio, $\hat{k}^*$. Equations (2.24), (2.25), and (2.26) determine the path of $\hat{k}$ and $\hat{c}$ for a given value of $\hat{k}(0)$. The phase diagram in figure 2.1 shows the nature of the dynamics.[19]

We first display the $\dot{\hat{c}} = 0$ locus. Since $\dot{\hat{c}} = \hat{c} \cdot (1/\theta) \cdot [f'(\hat{k}) - \delta - \rho - \theta x]$, there are two ways for $\dot{\hat{c}}$ to be zero: $\hat{c} = 0$, which corresponds to the horizontal axis in figure 2.1, and $f'(\hat{k}) = \delta + \rho + \theta x$, which is a vertical line at $\hat{k}^*$, the capital-labor ratio defined in equation (2.29). We note that $\hat{c}$ is rising for $\hat{k} < \hat{k}^*$ (so the arrows point upward in this region) and falling for $\hat{k} > \hat{k}^*$ (where the arrows point downward).

Recall that the solid curve in figure 2.1 shows combinations of $\hat{k}$ and $\hat{c}$ that satisfy $\dot{\hat{k}} = 0$ in equation (2.24). This equation also implies that $\hat{k}$ is falling for values of $\hat{c}$ above the solid curve (so the arrows point leftward in this region) and rising for values of $\hat{c}$ below the curve (where the arrows point rightward).

Since the $\dot{\hat{c}} = 0$ and the $\dot{\hat{k}} = 0$ loci cross three times, there are three steady states: the first one is the origin ($\hat{c} = \hat{k} = 0$), the second steady state corresponds to $\hat{k}^*$ and $\hat{c}^*$, and

19. See the appendix on mathematics for a discussion of phase diagrams.

the third one involves a positive capital stock, $\hat{k}^{**} > 0$, but zero consumption. We neglect the solution at the origin because it is uninteresting.

The second steady state is saddle-path stable. Note, in particular, that the pattern of arrows in figure 2.1 is such that the economy can converge to this steady state if it starts in two of the four quadrants in which the two schedules divide the space. The saddle-path property can also be verified by linearizing the system of dynamic equations around the steady state and noting that the determinant of the characteristic matrix is negative (see appendix 2A, section 2.8, for details). This sign for the determinant implies that the two eigenvalues have opposite signs, an indication that the system is locally saddle-path stable.

The dynamic equilibrium follows the stable saddle path shown by the solid locus with arrows. Suppose, for example, that the initial factor ratio satisfies $\hat{k}(0) < \hat{k}^*$, as shown in figure 2.1. If the initial consumption ratio is $\hat{c}(0)$, as shown, the economy follows the stable path toward the steady-state pair, $(\hat{k}^*, \hat{c}^*)$. This path satisfies all the first-order conditions, including the transversality condition, as shown in the previous section.

The two other possibilities are that the initial consumption ratio exceeds or falls short of $\hat{c}(0)$. If the ratio exceeds $\hat{c}(0)$, the initial saving rate is too low for the economy to remain on the stable path. The trajectory eventually crosses the $\dot{\hat{k}} = 0$ locus. After that crossing, $\hat{c}$ continues to rise, $\hat{k}$ starts to decline, and the path hits the vertical axis in finite time, at which point $\hat{k} = 0$.[20] The condition $f(0) = 0$ implies $\hat{y} = 0$; therefore, $\hat{c}$ must jump downward to 0 at this point. Because this jump violates the first-order condition that underlies equation (2.25), these paths—in which the initial consumption ratio exceeds $\hat{c}(0)$—are not equilibria.[21]

The final possibility is that the initial consumption ratio is below $\hat{c}(0)$. In this case, the initial saving rate is too high to remain on the saddle path, and the economy eventually crosses the $\dot{\hat{c}} = 0$ locus. After that crossing, $\hat{c}$ declines and $\hat{k}$ continues to rise. The economy converges to the point at which the $\dot{\hat{k}} = 0$ schedule intersects the horizontal axis, a point which we labeled $\hat{k}^{**}$. Note, in particular, that $\hat{k}$ rises above the golden-rule value, $\hat{k}_{\text{gold}}$, and asymptotically approaches a higher value of $\hat{k}$. Therefore, $f'(\hat{k}) - \delta$ falls below $x + n$ asymptotically, and the path violates the transversality condition given in equation (2.26). This violation of the transversality condition means that households are oversaving: utility

20. We can verify from equation (2.24) that $\dot{\hat{k}}$ becomes more and more negative in this region. Therefore, $\hat{k}$ must reach 0 in finite time.

21. This analysis applies if investment is reversible. If investment is irreversible, the constraint $\hat{c} \leq f(\hat{k})$ becomes binding before the trajectory hits the vertical axis. That is, the paths that start from points such as $\hat{c}_0'$ in figure 2.1 would eventually hit the production function, $\hat{c} = f(\hat{k})$, which lies above the locus for $\dot{\hat{k}} = 0$. Thereafter, the path would follow the production function downward toward the origin. Appendix 2B (section 2.9) shows that such paths are not equilibria.

would increase if consumption were raised at earlier dates. Accordingly, paths in which the initial consumption ratio is below $\hat{c}(0)$ are not equilibria. This result leaves the stable saddle path leading to the positive steady state, $\hat{k}^*$, as the only possibility.[22]

2.6.2 The Importance of the Transversality Condition

It is important to emphasize the role of the transversality condition in the determination of the unique equilibrium. To make this point, we consider an unrealistic variant of the Ramsey model in which everyone knows that the world will end at some known date $T > 0$. The utility function in equation (2.1) then becomes

$$U = \int_0^T u[c(t)] \cdot e^{nt} \cdot e^{-\rho t}\, dt$$

and the non-Ponzi condition is

$$a(T) \cdot \exp\left[-\int_0^T [r(v) - n]\, dv\right] \geq 0$$

The budget constraint is still given by equation (2.3). Since the only difference between this problem and that of the previous sections is the terminal date, the only optimization condition that changes is the transversality condition, which is now

$$a(T) \cdot \exp\left[-\int_0^T [r(v) - n]\, dv\right] = 0$$

Since the exponential term cannot be zero in finite time, this condition implies that the assets left at the end of the planning horizon equal zero:

$$a(T) = 0 \tag{2.32}$$

In other words, since the shadow value of assets at time T is positive, households will optimally choose to leave no assets when they "die."

The behavior of firms is the same as before, and equilibrium in the asset markets again requires $a(t) = k(t)$. Therefore, the general-equilibrium conditions are still given by equations (2.24) and (2.25), and the loci for $\dot{\hat{k}} = 0$ and $\dot{\hat{c}} = 0$ are the same as those shown

22. Similar results apply if the economy begins with $\hat{k}(0) > \hat{k}^*$ in figure 2.1. The only complication here is that, if investment is irreversible, the constraint $\hat{c} \leq f(\hat{k})$ may be binding in this region. See the discussion in appendix 2B (section 2.9).

in figure 2.1. The arrows representing the dynamics of the system are also the same as before.

Since $a(t) = k(t)$, the transversality condition from equation (2.32) can be written as

$$\hat{k}(T) = 0 \tag{2.33}$$

From the perspective of figure 2.1, this new transversality condition requires the initial choice of $\hat{c}(0)$ to be such that the capital stock equals zero at time T. In other words, optimality now requires the economy to land on the vertical axis at exactly time T. The implication is that the stable arm is no longer the equilibrium, because it is does not lead the economy toward zero capital at time T. The same is true for any initial choice of consumption below the stable arm. The new equilibrium, therefore, features an initial value $\hat{c}(0)$ that lies above the stable arm.

It is possible that $\hat{c}$ and $\hat{k}$ would both rise for awhile. In fact, if T is large, the transition path would initially be close to, but slightly above, the stable arm shown in figure 2.1. However, the economy eventually crosses the $\dot{\hat{k}} = 0$ schedule. Thereafter, $\hat{c}$ and $\hat{k}$ fall, and the economy ends up with zero capital at time T. We see, therefore, that the same system of differential equations involves one equilibrium (the stable arm) or another (the path that ends up on the vertical axis at T) depending solely on the transversality condition.

2.6.3 The Shape of the Stable Arm

The stable arm shown in figure 2.1 expresses the equilibrium $\hat{c}$ as a function of $\hat{k}$.[23] This relation is known in dynamic programming as a *policy function:* it relates the optimal value of a control variable, $\hat{c}$, to the state variable, $\hat{k}$. This policy function is an upward-sloping curve that goes through the origin and the steady-state position. Its exact shape depends on the parameters of the model.

Consider, as an example, the effect of the parameter θ on the shape of the stable arm. Suppose that the economy begins with $\hat{k}(0) < \hat{k}^*$, so that future values of $\hat{c}$ will exceed $\hat{c}(0)$. High values of θ imply that households have a strong preference for smoothing consumption over time; hence, they will try hard to shift consumption from the future to the present. Therefore, when θ is high, the stable arm will lie close to the $\dot{\hat{k}} = 0$ schedule, as shown in figure 2.2. The correspondingly low rate of investment implies that the transition would take a long time.

Conversely, if θ is low, households are more willing to postpone consumption in response to high rates of return. The stable arm in this case is flat and close to the horizontal axis for

23. The corresponding relation in the Solow–Swan model, $\hat{c} = (1 - s) \cdot f(\hat{k})$, was provided by the assumption of a constant saving rate.

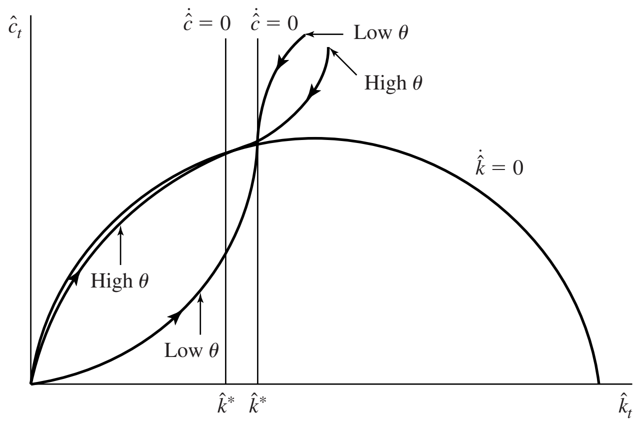

Figure 2.2
The slope of the saddle path. When θ is low, consumers do not mind large swings in consumption over time. Hence, they choose to consume relatively little when the capital stock is low (and the interest rate is high). The investment rate is high initially in this situation, and the economy approaches its steady state rapidly. In contrast, when θ is high, consumers are strongly motivated to smooth consumption over time. Hence, they initially devote most of their resources to consumption (the stable arm is close to the $\dot{\hat{k}} = 0$ schedule) and little to investment. In this case, the economy approaches its steady state slowly.

low values of $\hat{k}$ (see figure 2.2). The high levels of investment imply that the transition is relatively quick, and as $\hat{k}$ approaches $\hat{k}^*$, households increase $\hat{c}$ sharply. It is clear from the diagram that linear approximations around the steady state will not capture these dynamics accurately.

We show in appendix 2C (section 2.10) for the case of a Cobb–Douglas technology, $\hat{y} = A\hat{k}^\alpha$, that $\hat{c}/\hat{k}$ is rising, constant, or falling in the transition from $\hat{k}(0) < \hat{k}^*$ depending on whether the parameter θ is smaller than, equal to, or larger than the capital share, α. It follows that the stable arm is convex, linear, or concave depending on whether θ is smaller than, equal to, or larger than α. (We argue later that $\theta > \alpha$ is the plausible case.) If $\theta = \alpha$, so that $\hat{c}/\hat{k}$ is constant during the transition, the policy function has the closed-form solution $\hat{c} = (\text{constant}) \cdot \hat{k}$, where the constant turns out to be $(\delta + \rho)/\theta - (\delta + n)$.

2.6.4 Behavior of the Saving Rate

The gross saving rate, s, equals $1 - \hat{c}/f(\hat{k})$. The Solow–Swan model, discussed in chapter 1, assumed that s was constant at an arbitrary level. In the Ramsey model with optimizing consumers, s can follow a complicated path that includes rising and falling segments as the economy develops and approaches the steady state.

Heuristically, the behavior of the saving rate is ambiguous because it involves the offsetting impacts from a substitution effect and an income effect. As $\hat{k}$ rises, the decline in $f'(\hat{k})$ lowers the rate of return, r, on saving. The reduced incentive to save—an intertemporal-substitution effect—tends to lower the saving rate as the economy develops. Second, the income per effective worker in a poor economy, $f(\hat{k})$, is far below the long-run or permanent income of this economy. Since households desire to smooth consumption, they would like to consume a lot in relation to income when they are poor; that is, the saving rate would be low when $\hat{k}$ is low. As $\hat{k}$ rises, the gap between current and permanent income diminishes; hence, consumption tends to fall in relation to current income, and the saving rate tends to rise. This force—an income effect—tends to raise the saving rate as the economy develops.

The transitional behavior of the saving rate depends on whether the substitution or income effect is more important. The net effect is ambiguous in general, and the path of the saving rate during the transition can be complicated. The results simplify, however, for a Cobb–Douglas production function. Appendix 2C shows for this case that, depending on parameter values, the saving rate falls monotonically, stays constant, or rises monotonically as $\hat{k}$ rises.

We show in Appendix 2C for the Cobb–Douglas case that the steady-state saving rate, s^*, is given by

$$s^* = \alpha \cdot (x + n + \delta)/(\delta + \rho + \theta x) \tag{2.34}$$

Note that the transversality condition, which led to equation (2.31), implies $s^* < \alpha$ in equation (2.34); that is, the steady-state gross saving rate is less than the gross capital share.

We can use a phase diagram to analyze the transitional behavior of the saving rate for the case of a Cobb–Douglas production function. The methodology is interesting more generally because it provides a way to study the behavior of variables of interest, such as the saving rate, that do not enter directly into the first-order conditions of the model. The method involves transformations of the variables that appear in the first-order conditions. The dynamic relations that we used before were written in terms of the variables $\hat{c}$ and $\hat{k}$. To study the transitional behavior of the saving rate, $s = 1 - \hat{c}/\hat{y}$, we want to rewrite these relations in terms of the variables $\hat{c}/\hat{y}$ and $\hat{k}$. Then we will be able to construct a phase diagram in terms of $\hat{c}/\hat{y}$ and $\hat{k}$. The stable arm of such a phase diagram will show how $\hat{c}/\hat{y}$—and, hence, $s = 1 - \hat{c}/\hat{y}$—move as $\hat{k}$ increases.

We start by noticing that the growth rate of $\hat{c}/\hat{y}$ is given by the growth rate of $\hat{c}$ minus the growth rate of $\hat{y}$. If the production function is Cobb–Douglas, the growth rate of $\hat{y}$ is proportional to the growth rate of $\hat{k}$, that is,

$$\frac{1}{\hat{c}/\hat{y}} \cdot \frac{d(\hat{c}/\hat{y})}{dt} = (\dot{\hat{c}}/\hat{c}) - (\dot{\hat{y}}/\hat{y}) = (\dot{\hat{c}}/\hat{c}) - \alpha \cdot (\dot{\hat{k}}/\hat{k})$$

We can now use the equilibrium conditions shown in equations (2.24) and (2.25) to get

$$\frac{1}{\hat{c}/\hat{y}} \cdot \frac{d(\hat{c}/\hat{y})}{dt} = [(1/\theta) \cdot (\alpha A\hat{k}^{\alpha-1} - \delta - \rho - \theta x)] - \alpha \cdot [A\hat{k}^{\alpha-1} - (\hat{c}/\hat{y}) \cdot A\hat{k}^{\alpha-1} - (x+n+\delta)] \tag{2.35}$$

where we used the equality $\hat{c}/\hat{k} = (\hat{c}/\hat{y}) \cdot A\hat{k}^{\alpha-1}$. The growth rate of $\hat{k}$ is

$$\dot{\hat{k}}/\hat{k} = [A\hat{k}^{\alpha-1} - (\hat{c}/\hat{y}) \cdot A\hat{k}^{\alpha-1} - (x+n+\delta)] \tag{2.36}$$

Notice that equations (2.35) and (2.36) represent a system of differential equations in the variables $\hat{c}/\hat{y}$ and $\hat{k}$. Therefore, a conventional phase diagram can be drawn in terms of these two variables.

We start by setting equation (2.35) to zero to get the $\frac{d(\hat{c}/\hat{y})}{dt} = 0$ locus:

$$\hat{c}/\hat{y} = \left(1 - \frac{1}{\theta}\right) + \psi \cdot \frac{\hat{k}^{1-\alpha}}{\alpha A} \tag{2.37}$$

where $\psi \equiv [(\delta + \rho + \theta x)/\theta - \alpha \cdot (x+n+\delta)]$ is a constant. This locus is upward sloping, downward sloping, or horizontal depending on whether ψ is positive, negative, or zero. The three possibilities are depicted in figure 2.3.

Independently of the value of ψ, the arrows above the $\frac{d(\hat{c}/\hat{y})}{dt} = 0$ locus point north, and the arrows below the schedule point south.

We can find the $\dot{\hat{k}} = 0$ locus by setting equation (2.35) to zero to get

$$\hat{c}/\hat{y} = 1 - \frac{(x+n+\delta)}{A} \cdot \hat{k}^{1-\alpha} \tag{2.38}$$

which is unambiguously downward sloping.[24] Arrows point west above the schedule and east below the schedule.

The three panels of figure 2.3 show that the steady state is saddle-path stable regardless of the value of ψ. The stable arm, however, is upward-sloping when $\psi > 0$, downward-sloping when $\psi < 0$, and horizontal when $\psi = 0$. Following the reasoning of previous sections, we know that an infinite-horizon economy always finds itself on the stable arm. Thus, depending on parameter values, the consumption ratio falls monotonically, stays constant, or rises monotonically as $\hat{k}$ rises. The saving rate, therefore, behaves exactly the opposite. A high value of θ—which corresponds to a low willingness to substitute consumption intertemporally—makes it more likely that $\psi < 0$ will hold, in which case the saving rate

24. When $\psi < 0$, the $\frac{d\hat{k}}{dt} = 0$ locus is also steeper than the $\frac{d(\hat{c}/\hat{y})}{dt} = 0$ schedule.

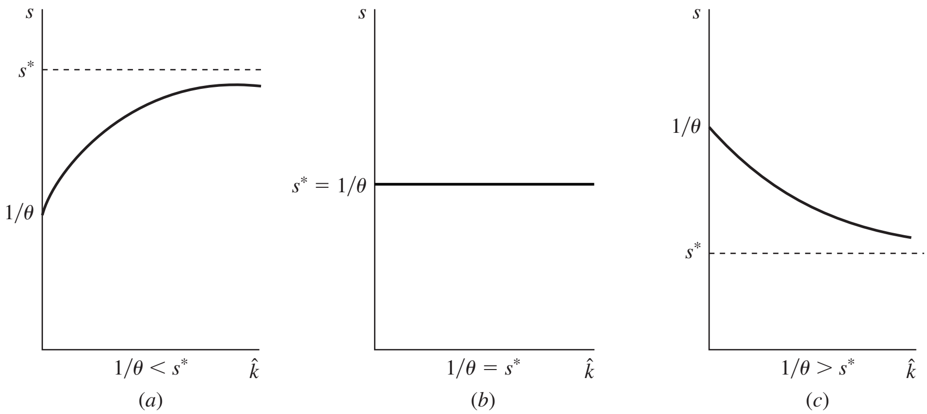

Figure 2.3
Phase diagram for the behavior of the saving rate (in the Cobb–Douglas case). In the Cobb–Douglas case, the savings rate behaves monotonically. Panel *a* shows the phase diagram for $\hat{c}/\hat{y}$ and $\hat{k}$ when the parameters are such that $(\delta + \rho + \theta x)/\theta > \alpha \cdot (x + n + \delta)$. Since the stable arm is upward sloping, the consumption ratio increases as the economy grows toward the steady state. Hence, in this case, the saving rate (one minus the consumption rate) declines monotonically during the transition. Panel *b* considers the case in which $(\delta + \rho + \theta x)/\theta < \alpha \cdot (x + n + \delta)$. The stable arm is now downward sloping and, therefore, the saving rate increases monotonically during the transition. Panel *c* considers the case $(\delta + \rho + \theta x)/\theta + \alpha \cdot (x + n + \delta)$. The stable arm is now horizonal, which means that the saving rate is constant during the transition.

is more likely to rise during the transition. This result follows because a higher θ weakens the substitution effect from the interest rate.

In the particular case where $\psi = 0$, the saving rate is constant at its steady-state value, $s^* = 1/\theta$, during the transition. For this combination of parameters, it turns out that the wealth and substitution effects cancel out, so that the saving rate remains constant as the capital stock grows toward its steady state. Thus, the constant saving rate in the Solow–Swan model is a special case of the Ramsey model. However, even in this case, there is an important difference from the Solow–Swan model. The level of s in the Ramsey model is dictated by the underlying parameters and cannot be chosen arbitrarily. In particular, an arbitrary choice of s in the Solow–Swan model may generate results that are dynamically inefficient if s leads the economy to a steady-state capital stock that is larger than the golden rule. This outcome is impossible in the Ramsey model.

In a later discussion, we use the baseline values $\rho = 0.02$ per year, $\delta = 0.05$ per year, $n = 0.01$ per year, and $x = 0.02$ per year. If we also assume a conventional capital share of $\alpha = 0.3$, the value of θ that generates a constant saving rate is 17; that is, $s^* < 1/\theta$ applies and the saving rate falls—counterfactually—as the economy develops unless θ exceeds this high value.

We noted for the Solow–Swan model that the theory cannot fit the evidence about speeds of convergence unless the capital-share coefficient, α, is much larger than 0.3. Values in the neighborhood of 0.75 accord better with the empirical evidence, and these high values of α are reasonable if we take a broad view of capital to include the human components. We show in the following section that the findings about α still apply in the Ramsey growth model, which allows the saving rate to vary over time. If we assume $\alpha = 0.75$, along with the benchmark values of the other parameters, the value of θ that generates a constant saving rate is 1.75. That is, the gross saving rate rises (or falls) as the economy develops if θ is greater (or less) than 1.75. If $\theta = 1.75$, the gross saving rate is constant at the value 0.57. We have to interpret this high value for the gross saving rate by including in gross saving the various expenditures that expand or maintain human capital; aside from expenses for education and training, this gross saving would include portions of the outlays for food, health, and so on.

Our reading of empirical evidence across countries is that the saving rate tends to rise to a moderate extent with per capita income during the transition. The Ramsey model can fit this pattern, as well as the observed speeds of convergence, if we combine the benchmark parameters with a value of α of around 0.75 and a value of θ somewhat above 2. The value of θ cannot be too much above 2 because then the steady-state saving rate, s^*, shown in equation (2.34), becomes too low. For example, the value $\theta = 10$ implies $s^* = 0.22$, which is too low for a broad concept that includes gross saving in the form of human capital.

2.6.5 The Paths of the Capital Stock and Output

The stable arm shown in figure 2.1 shows that, if $\hat{k}(0) < \hat{k}^*$, $\hat{k}$ and $\hat{c}$ rise monotonically as they approach their steady-state values. The rising path of $\hat{k}$ implies that the rate of return, r, declines monotonically from its initial position, $f'[\hat{k}(0)] - \delta$, to its steady-state value, $\rho + \theta x$. Equation (2.25) and the path of decreasing r imply that the growth rate of per capita consumption, $\dot{c}/c$, falls monotonically. That is, the lower $\hat{k}(0)$ and, hence, $\hat{y}(0)$, the higher the initial value of $\dot{c}/c$.

We would also like to relate the initial per capita growth rates of capital and output, γ_k and γ_y, to the starting ratio, $\hat{k}(0)$. In chapter 1 we referred to the negative relations between $\dot{k}/k$ and $\hat{k}(0)$ and between $\dot{y}/y$ and $\hat{y}(0)$ as convergence effects. We show in appendix 2D (section 2.11), using the consumption function from equations (2.15) and (2.16), that $\dot{k}/k$ declines monotonically as the economy develops and approaches the steady state. In other words, although the saving rate may rise during the transition, it cannot rise enough to eliminate the inverse relation between $\dot{k}/k$ and $\hat{k}$. Thus, the endogenous determination of the saving rate does not eliminate the convergence property for $\hat{k}$.

We can take logs and derivatives of the production function in equation (2.18) to derive the growth rate of output per effective worker:

$$\dot{\hat{y}}/\hat{y} = \left[\frac{\hat{k} \cdot f'(\hat{k})}{f(\hat{k})}\right] \cdot (\dot{\hat{k}}/\hat{k}) \tag{2.39}$$

that is, the growth rate of $\hat{k}$ is multiplied by the share of gross capital income in gross product. For a Cobb–Douglas production function, the share of capital income equals the constant α. Therefore, the properties of $\dot{k}/k$ carry over immediately to those of $\dot{y}/y$. This result applies more generally than in the Cobb–Douglas case unless the share of capital income rises fast enough as an economy develops to more than offset the fall in $\dot{k}/k$.

2.6.6 Speeds of Convergence

Log-Linear Approximations Around the Steady State We want now to provide a quantitative assessment of the speed of convergence in the Ramsey model. We begin with a log-linearized version of the dynamic system for $\hat{k}$ and $\hat{c}$, equations (2.24) and (2.25). This approach is an extension of the method that we used in chapter 1 for the Solow–Swan model; the only difference here is that we have to deal with a two-variable system instead of a one-variable system. The advantage of the log-linearization method is that it provides a closed-form solution for the convergence coefficient. The disadvantage is that it applies only as an approximation in the neighborhood of the steady state.

Appendix 2A examines a log-linearized version of equations (2.24) and (2.25) when expanded around the steady-state position. The results can be written as

$$\log[\hat{y}(t)] = e^{-\beta t} \cdot \log[\hat{y}(0)] + (1 - e^{-\beta t}) \cdot \log(\hat{y}^*) \tag{2.40}$$

where $\beta > 0$. Thus, for any $t \geq 0$, $\log[\hat{y}(t)]$ is a weighted average of the initial and steady-state values, $\log[\hat{y}(0)]$ and $\log(\hat{y}^*)$, with the weight on the initial value declining exponentially at the rate β. The speed of convergence, β, depends on the parameters of technology and preferences. For the case of a Cobb–Douglas technology, the formula for the convergence coefficient (which comes from the log-linearization around the steady-state position) is

$$2\beta = \left\{\zeta^2 + 4 \cdot \left(\frac{1-\alpha}{\theta}\right) \cdot (\rho + \delta + \theta x) \cdot \left[\frac{\rho + \delta + \theta x}{\alpha} - (n + x + \delta)\right]\right\}^{1/2} - \zeta \tag{2.41}$$

where $\zeta = \rho - n - (1 - \theta) \cdot x > 0$. We discuss below the way that the various parameters enter into this formula.

Equation (2.40) implies that the average growth rate of per capita output, y, over an interval from an initial time 0 to any future time $T \geq 0$ is given by

$$(1/T) \cdot \log[y(T)/y(0)] = x + \frac{(1 - e^{-\beta T})}{T} \cdot \log[\hat{y}^*/\hat{y}(0)] \tag{2.42}$$

Hold fixed, for the moment, the steady-state growth rate x, the convergence speed β, and the averaging interval T. Then equation (2.42) says that the average per capita growth rate of output depends negatively on the ratio of $\hat{y}(0)$ to $\hat{y}^*$. Thus, as in the Solow–Swan model, the effect of the initial position, $\hat{y}(0)$, is conditioned on the steady-state position, $\hat{y}^*$. In other words, the Ramsey model also predicts conditional, rather than absolute, convergence.

The coefficient that relates the growth rate of y to $\log[\hat{y}^*/\hat{y}(0)]$ in equation (2.42), $(1 - e^{-\beta T})/T$, declines with T for given β. If $\hat{y}(0) < \hat{y}^*$, so that growth rates decline over time, an increase in T means that more of the lower future growth rates are averaged with the higher near-term growth rates. Therefore, the average growth rate, which enters into equation (2.42), falls as T rises. As $T \to \infty$, the steady-state growth rate, x, dominates the average; hence, the coefficient $(1 - e^{-\beta T})/T$ approaches 0, and the average growth rate of y in equation (2.42) tends to x.

For a given T, a higher β implies a higher coefficient $(1 - e^{-\beta T})/T$. (As $T \to 0$, the coefficient approaches β.) Equation (2.41) expresses the dependence of β on the underlying parameters. Consider first the case of the Solow–Swan model in which the saving rate is constant. As noted before, this situation applies if the steady-state saving rate, s^*, shown in equation (2.34) equals $1/\theta$ or, equivalently, if the combination of parameters $\alpha \cdot (\delta + n) - (\delta + \rho)/\theta - x \cdot (1 - \alpha)$ equals 0.

Suppose that the parameters take on the baseline values that we used in chapter 1: $\delta = 0.05$ per year, $n = 0.01$ per year, and $x = 0.02$ per year. We also assume $\rho = 0.02$ per year to get a reasonable value for the steady-state interest rate, $\rho + \theta x$. As mentioned in a previous section, for these benchmark parameter values, the saving rate is constant if $\alpha = 0.3$ when $\theta = 17$ and if $\alpha = 0.75$ when $\theta = 1.75$.

With a constant saving rate, the formula for the convergence speed, β, simplifies from equation (2.41) to the result that applied in equation (1.45) for the Solow–Swan model:

$$\beta^* = (1 - \alpha) \cdot (x + n + \delta)$$

We noted in chapter 1 that a match with the empirical estimate for β of roughly 0.02 per year requires a value for α around 0.75, that is, in the range in which the broad nature of capital implies that diminishing returns to capital set in slowly. Lower values of $x + n + \delta$ reduce the required value of α, but plausible values leave α well above the value of around 0.3, which would apply to a narrow concept of physical capital.

In the case of a variable saving rate, equation (2.41) determines the full effects of the various parameters on the convergence speed. The new element concerns the tilt of the time path of the saving rate during the transition. If the saving rate falls with $\hat{k}$, the convergence speed would be higher than otherwise, and vice versa. For example, we found before that a higher value of the intertemporal-substitution parameter, θ, makes it more likely that the saving rate would rise with $\hat{k}$. Through this mechanism, a higher θ reduces the speed of convergence, β, in equation (2.41).

If the rate of time preference, ρ, increases, the level of the saving rate tends to fall (see equation [2.34]). The effect on the convergence speed depends, however, not on the level of the saving rate but on the tendency for the saving rate to rise or fall as the economy develops. A higher ρ tends to tilt downward the path of the saving rate. The effective time-preference rate is $\rho + \theta \cdot \dot{c}/c$. Because $\dot{c}/c$ is inversely related to $\hat{k}$, the impact of ρ on the effective time-preference rate is proportionately less the lower is $\hat{k}$. Therefore, the saving rate tends to decrease less the lower $\hat{k}$, and, hence, the time path of the saving rate tilts downward. A higher ρ tends accordingly to raise the magnitude of β in equation (2.41).

It turns out with a variable saving rate that the parameters δ and x tend to raise β, just as they did in the Solow–Swan model. The overall effect from the parameter n becomes ambiguous but tends to be small in the relevant range.[25]

The basic result, which holds with a variable or constant saving rate, is that, for plausible values of the other parameters, the model requires a high value of α—in the neighborhood of 0.75—to match empirical estimates of the speed of convergence, β. We can reduce the required value of α to 0.5–0.6 if we assume very high values of θ (in excess of 10) along with a value of δ close to 0. We argued before, however, that very high values of θ make the steady-state saving rate too low, and values of δ near 0 are unrealistic. In addition, as we show later, values of α that are much below 0.75 generate counterfactual predictions about the transitional behavior of the interest rate and the capital-output ratio. We discuss in chapter 3 how adjustment costs for investment can slow down the rate of convergence, but this extension does not change the main conclusions.

Numerical Solutions of the Nonlinear System We now assess the convergence properties of the model with a second approach, which uses numerical methods to solve the nonlinear system of differential equations. This approach avoids the approximation errors inherent in linearization of the model and provides accurate results for a given specification of the underlying parameters. The disadvantage is the absence of a closed-form solution. We have to generate a new set of answers for each specification of parameter values.

25. Equation (2.41) implies that the effects on β are unambiguously negative for α and positive for δ. Our numerical computations indicate that the effects of the other parameters are in the directions that we mentioned as long as the other parameters are restricted to a reasonable range.

We can use numerical methods to obtain a global solution for the nonlinear system of differential equations. In the case of a Cobb–Douglas production function, the growth rates of $\hat{k}$ and $\hat{c}$ are given from equations (2.24) and (2.25) as

$$\gamma_{\hat{k}} \equiv \dot{\hat{k}}/\hat{k} = A \cdot (\hat{k})^{\alpha-1} - (\hat{c}/\hat{k}) - (x + n + \delta) \tag{2.43}$$

$$\gamma_{\hat{c}} \equiv \dot{\hat{c}}/\hat{c} = (1/\theta) \cdot [\alpha A \cdot (\hat{k})^{\alpha-1} - (\delta + \rho + \theta x)] \tag{2.44}$$

If we specified the values of the parameters $(A, \alpha, x, n, \delta, \rho, \theta)$ and knew the relation between $\hat{c}$ and $\hat{k}$ along the path—that is, if we knew the policy function $\hat{c}(\hat{k})$—then standard numerical methods for solving differential equations would allow us to solve out for the entire time paths of $\hat{k}$ and $\hat{c}$. The appendix on mathematics shows how to use a procedure called the *time-elimination method* to derive the policy function numerically. (See also Mulligan and Sala-i-Martin, 1991). We assume now that we have already solved this part of the problem.

Once we know the policy function, we can determine the paths of all the variables that we care about, including the convergence coefficient, defined by $\beta = -\, d(\gamma_{\hat{k}})/d[\log(\hat{k})]$. (In the Cobb–Douglas case, the convergence coefficient for $\hat{y}$ is still the same as that for $\hat{k}$.) Figure 2.4 shows the relation between β and $\hat{k}/\hat{k}^*$ when we use our benchmark parameter values ($\delta = 0.05$, $x = 0.02$, $n = 0.01$, $\rho = 0.02$), $\theta = 3$, and $\alpha = 0.3$ or 0.75.[26] For either setting of α, β is a decreasing function of $\hat{k}/\hat{k}^*$; that is, the speed of convergence slows down as the economy approaches the steady state.[27] At the steady state, where $\hat{k}/\hat{k}^* = 1$, the values of β—0.082 if $\alpha = 0.3$ and 0.015 if $\alpha = 0.75$—are those implied by equation (2.41) for the log-linearization around the steady state.

If $\hat{k}/\hat{k}^* < 1$, figure 2.4 indicates that β exceeds the values implied by equation (2.41). For example, if $\hat{k}/\hat{k}^* = 0.5$, $\beta = 0.141$ if $\alpha = 0.3$ and 0.018 if $\alpha = 0.75$. If $\hat{k}/\hat{k}^* = 0.1$, $\beta = 0.474$ if $\alpha = 0.3$ and 0.026 if $\alpha = 0.75$. Thus, if we use our preferred high value for the capital-share coefficient, $\alpha = 0.75$, the convergence coefficient, β, remains between 1.5 percent and 3 percent for a broad range of $\hat{k}/\hat{k}^*$. This behavior accords with the empirical evidence discussed in chapters 11 and 12; we find there that convergence coefficients do not seem to exceed this range even for economies that are very far from their steady states. In contrast, if we assume $\alpha = 0.3$, the model incorrectly predicts extremely high rates of convergence when $\hat{k}$ is far below $\hat{k}^*$.

Since the convergence speeds rise with the distance from the steady state, the durations of the transition are shorter than those implied by the linearized model. We can use the results on the time path of $\hat{k}$ to compute the exact time that it takes to close a specified percentage

26. For a given value of $\hat{k}/\hat{k}^*$, the parameter A does not affect β in the Cobb–Douglas case.

27. This relation does not hold in general. In particular, β can rise with $\hat{k}/\hat{k}^*$ if θ is very small and α is very large, for example, if $\theta = 0.5$ and $\alpha = 0.95$.

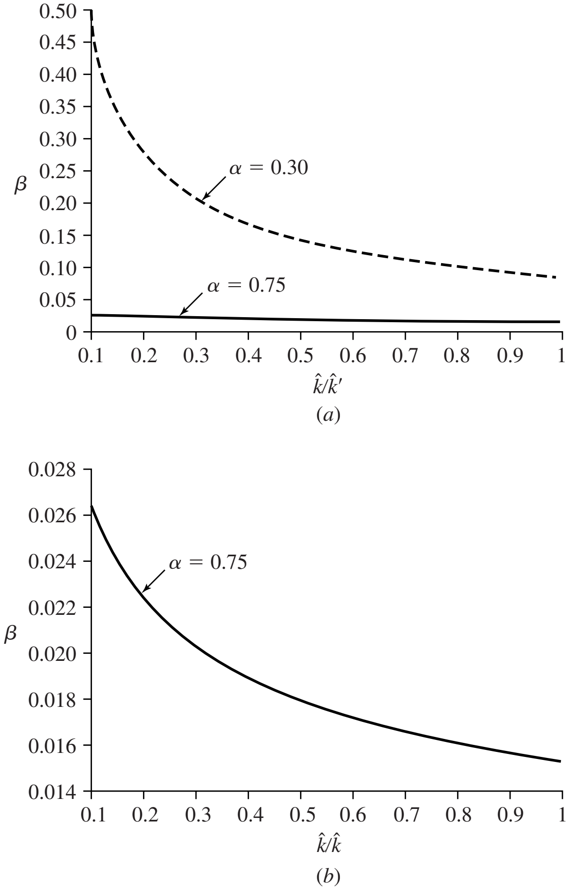

Figure 2.4
Numerical estimates of the speed of convergence in the Ramsey model. The exact speed of convergence (displayed on the vertical axis) is a decreasing function of the distance from the steady state, $\hat{k}/\hat{k}^*$ (shown on the horizontal axis). The analysis assumes a Cobb–Douglas production function, with results reported for two values of the capital share, $\alpha = 0.30$ and $\alpha = 0.75$. The change in the convergence speed during the transition is more pronounced for the smaller capital share. The value of the convergence speed, β, at the steady state ($\hat{k}/\hat{k}^* = 1$) is the value that we found analytically with a log-linear approximation around the steady state (equation [2.41]).

of the initial gap from $\hat{k}^*$. Panel a of figure 2.5 shows how the gap between $\hat{k}$ and $\hat{k}^*$ is eliminated over time if the economy begins with $\hat{k}/\hat{k}^* = 0.1$ and if $\alpha = 0.3$ or 0.75. As an example, if $\alpha = 0.75$, it takes 38 years to close 50 percent of the gap, compared with 45 years from the linear approximation.

Panel b in figure 2.5 displays the level of consumption, expressed as $\hat{c}/\hat{c}^*$; panel c the level of output, $\hat{y}/\hat{y}^*$; and panel d the level of gross investment, $\hat{\imath}/\hat{\imath}^*$. Note that for $\alpha = 0.75$, the paths of $\hat{c}/\hat{c}^*$ and $\hat{y}/\hat{y}^*$ are similar, because the gross saving rate and, hence, $\hat{c}/\hat{y}$ change only by small amounts in this case (discussed later).

Panel e shows $\gamma_{\hat{y}}$, the growth rate of $\hat{y}$. For $\alpha = 0.3$, the model has the counterfactual implication that the initial value of $\gamma_{\hat{y}}$ (corresponding to $\hat{k}/\hat{k}^* = 0.1$) is implausibly large, about 15 percent per year, which means that γ_y is about 17 percent per year. This kind of result led King and Rebelo (1993) to dismiss the transitional behavior of the Ramsey model as a reasonable approximation to actual growth experiences. We see, however, that for $\alpha = 0.75$, the model predicts more reasonably that $\gamma_{\hat{y}}$ would begin at about 3.5 percent per year, so that γ_y would be about 5.5 percent per year.

Panel f shows the gross saving rate, $s(t)$. We know from our previous analytical results for the Cobb–Douglas case, given the assumed values of the other parameters, that $s(t)$ falls monotonically when $\alpha = 0.3$ and rises monotonically when $\alpha = 0.75$. For $\alpha = 0.3$, the results are counterfactual in that the model predicts a fall in $s(t)$ from 0.28 at $\hat{k}/\hat{k}^* = 0.1$ to 0.22 at $\hat{k}/\hat{k}^* = 0.5$ and 0.18 at $\hat{k}/\hat{k}^* = 1$. The predicted levels of the saving rate are also unrealistically low for a broad concept of capital. In contrast, for $\alpha = 0.75$, the moderate rise in the saving rate as the economy develops fits well with the data. The saving rate rises in this case from 0.41 at $\hat{k}/\hat{k}^* = 0.1$ to 0.44 at $\hat{k}/\hat{k}^* = 0.5$ and 0.46 at $\hat{k}/\hat{k}^* = 1$. The predicted level of the saving rate is also reasonable if we take a broad view of capital.

Panel g displays the behavior of the interest rate, r. Note that the steady-state interest rate is $r^* = \rho + \theta x = 0.08$, and the corresponding marginal product is $f'(\hat{k}^*) = r^* + \delta = 0.13$. If we consider the initial position $\hat{k}(0)/\hat{k}^* = 0.1$, as in figure 2.5, the Cobb–Douglas production function implies

$$f'[\hat{k}(0)]/f'(\hat{k}^*) = [\hat{k}(0)/\hat{k}^*]^{\alpha-1} = (10)^{1-\alpha}$$

Hence, for $\alpha = 0.3$, we get $f'[\hat{k}(0)] = 5 \cdot f'(\hat{k}^*) = 0.55$. In other words, with a capital-share coefficient of around 0.3, the initial interest rate (at $\hat{k}[0]/\hat{k}^* = 0.1$) would take on the unrealistically high value of 60 percent. This counterfactual prediction about interest rates was another consideration that led King and Rebelo (1993) to reject the transitional dynamics of the Ramsey model. However, if we assume our preferred capital-share coefficient, $\alpha = 0.75$, we get $f'[\hat{k}(0)] = 1.8 \cdot f'(\hat{k}^*) = 0.23$, so that $r(0)$ takes on the more reasonable value of 18 percent.

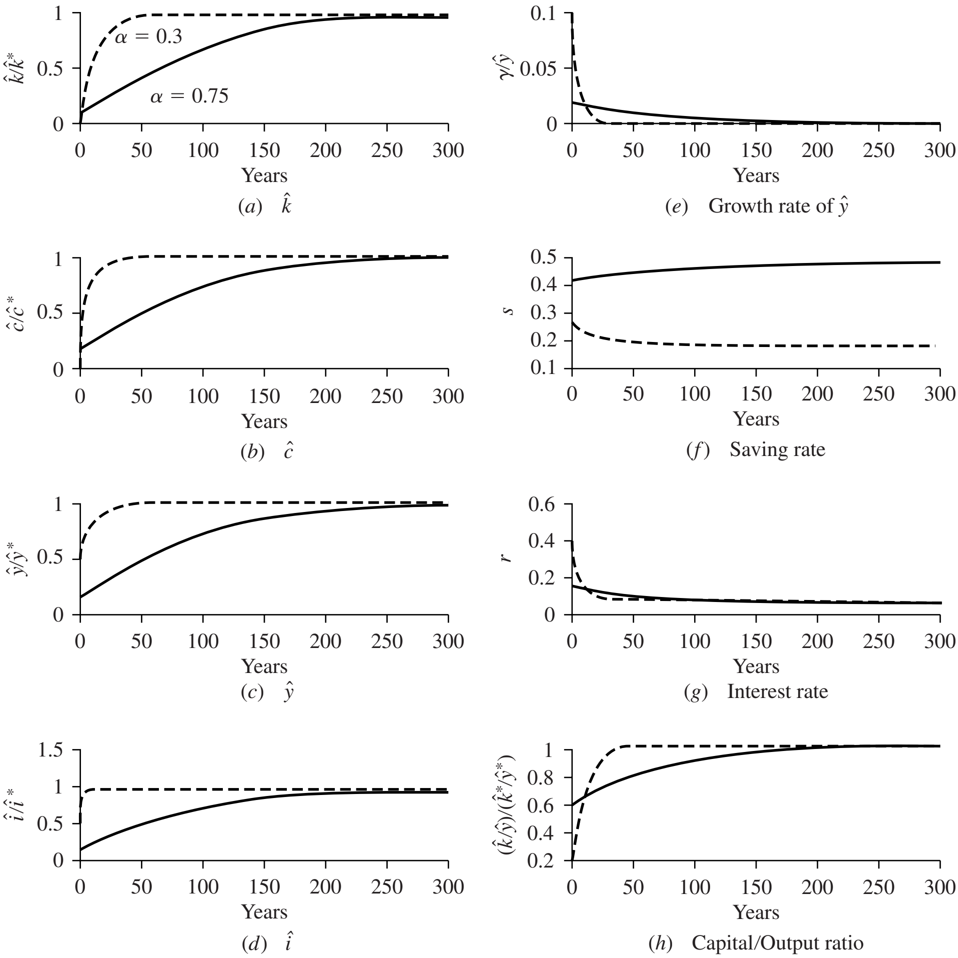

Figure 2.5

Numerical estimates of the dynamic paths in the Ramsey model. The eight panels display the exact dynamic paths of eight key variables: the values per unit of effective labor of the capital stock, consumption, output, and investment, the growth rate of output per effective worker, the saving rate, the interest rate, and the capital-output ratio. The first four variables and the last one are expressed as ratios to their steady-state values; hence, each variable approaches 1 asymptotically. The analysis assumes a Cobb–Douglas production technology, where the dotted line in each panel corresponds to $\alpha = 0.30$ and the solid line to $\alpha = 0.75$. The other parameters are reported in the text. The initial capital per effective worker is assumed in each case to be one-tenth of its steady-state value.

The final panel in figure 2.5 shows the behavior of the capital-output ratio, $(\hat{k}/\hat{y})$, expressed in relation to $(\hat{k}^*/\hat{y}^*)$. Kaldor (1963) argued that this ratio changed relatively little during the course of economic development, and Maddison (1982, chapter 3) supported this view. These observations pertain, however, to a narrow concept of physical capital, whereas our model takes a broad perspective to include human capital. The cross-country data show that places with higher real per capita GDP tend to have higher ratios of human capital in the form of educational attainment to physical capital (see Judson, 1998). This observation suggests that the ratio of human to physical capital would tend to rise during the transition to higher levels of real per capita GDP (see chapter 5 for a theoretical discussion of this behavior). If the ratio of physical capital to output remains relatively stable, the capital-output ratio for a broad measure of capital would increase during the transition.

With a Cobb–Douglas production function, the capital-output ratio is $\hat{k}/\hat{y} = (1/A) \cdot (\hat{k})^{(1-\alpha)}$. If $\alpha = 0.3$, an increase in $\hat{k}$ by a factor of 10 would raise $\hat{k}/\hat{y}$ by a factor of 5, a shift that departs significantly from the observed variations in $\hat{k}/\hat{y}$ over long periods of economic development. In contrast, if $\alpha = 0.75$, an increase in $\hat{k}$ by a factor of 10 would raise $\hat{k}/\hat{y}$ by a factor of only 1.8. For a broad concept of capital, this behavior appears reasonable.

The main lesson from the study of the time paths in figure 2.5 is that the transitional dynamics of the Ramsey model with a conventional capital-share coefficient, α, of around 0.3 does not provide a good description of various aspects of economic development. For an economy that starts far below its steady-state position, the inaccurate predictions include an excessive speed of convergence, unrealistically high growth and interest rates, a rapidly declining gross saving rate, and large increases over time in the capital-output ratio. All of these shortcomings are eliminated if we take a broad view of capital and assume a correspondingly high capital-share coefficient, α, of around 0.75. This value of α, together with plausible values of the model's other parameters, generate predictions that accord well with the growth experiences that we study in chapters 11 and 12.

2.6.7 Household Heterogeneity

Our analysis thus far has considered a single household as representing the entire economy. The consumption and saving decisions of the representative agent are supposed to capture the behavior of the average agent in a complex economy with many families. The important question is whether the behavior of this "representative" or "average" household is really equivalent to what we would get if we averaged the behavior of many heterogeneous families.

Caselli and Ventura (2000) have extended the Ramsey model to allow for various forms of household heterogeneity.[28] Following their analysis, we assume that the economy contains $\mathcal{J}$ equal-sized households, each of which is an infinitely lived dynasty. The population of each household—and, therefore, the overall population—grow at the constant rate n. Preferences of each household are still given by equations (2.1) and (2.10), with the preference parameters ρ and θ the same for each household. In this case, it is straightforward to allow for differences across households in initial assets and labor productivity.

Let $a_j(t)$ and π_j represent, respectively, the per capita assets and productivity level of the jth household. The wage rate paid to the jth household is $\pi_j w$, where w is the economy-wide average wage, π_j is constant over time, and we have normalized so that the mean value of π_j equals 1.

The flow budget constraint for each household takes the same form as equation (2.3):

$$\dot{a}_j = \pi_j \cdot w + ra_j - c_j - na_j \tag{2.45}$$

In this representation, each household could have a different value of initial assets, $a_j(0)$. The optimal growth rate of each household's per capita consumption satisfies the usual first-order condition from equation (2.9):

$$\dot{c}_j/c_j = (1/\theta) \cdot (r - \rho) \tag{2.46}$$

The household's level of per capita consumption can be found, as in the analysis of the first section of this chapter, by solving out the differential equation for c_j and using the transversality condition (of the form of equation [2.12]). The result, analogous to equation (2.15), is

$$c_j = \mu \cdot (a_j + \pi_j \tilde{w}) \tag{2.47}$$

where μ is the propensity to consume out of assets (given by equation [2.16]) and $\tilde{w}$ is the present value of the economy-wide average wage.

The economy-wide value of per capita assets is $a = (\frac{1}{\mathcal{J}}) \cdot \sum_1^{\mathcal{J}} a_j$, and the economy-wide value of per capita consumption is $c = (\frac{1}{\mathcal{J}}) \cdot \sum_1^{\mathcal{J}} c_j$. Since the population growth rate is the same for all households, aggregation is straightforward: sum equation (2.45) over the $\mathcal{J}$ households and divide by $\mathcal{J}$ to compute the economy-wide budget constraint:

$$\dot{a} = w + ra - c - na \tag{2.48}$$

This budget constraint is the same as equation (2.3).

28. Stiglitz (1969) worked out a model with household heterogeneity under a variety of nonoptimizing saving functions.

We can also aggregate the consumption function, equation (2.47), across households to get the economy-wide value of consumption per person:

$$c = \mu \cdot (a + \tilde{w}) \tag{2.49}$$

This relation is the same as equation (2.15).

Finally, we can use equations (2.48) and (2.49) to get

$$\dot{c}/c = (1/\theta) \cdot (r - \rho) \tag{2.50}$$

which is the standard economy-wide condition for consumption growth. When combined with the usual analysis of competitive firms, this description of aggregate household behavior—equations (2.48) and (2.50)—delivers the standard Ramsey model. Hence, the model with the assumed forms of heterogeneity in initial assets and worker productivity has the same macroeconomic implications as the usual, representative-agent model. In other words, if the households in the economy differ in their level of wealth or productivity, and if their preferences are CIES with identical parameters and discount rates, the average consumption, assets, income, and capital for these families behave exactly as the ones of a single representative household. Hence, the representative-agent model provides the correct description of the average variables of an economy populated with the assumed forms of heterogenous agents.

Aside from supporting the use of the representative-agent framework, the extension to include heterogeneity also allows for a study of the dynamics of inequality. Equation (2.46) implies that each household chooses the same growth rate for consumption. Therefore, relative consumption, c_j/c, does not vary over time.

The model does imply a dynamics for relative assets, a_j/a. Equations (2.45), (2.47), (2.48), and (2.49) imply that relative assets change in accordance with

$$\frac{d}{dt}\left(\frac{a_j}{a}\right) = \frac{(w - \mu\tilde{w})}{a} \cdot \left(\pi_j - \frac{a_j}{a}\right) \tag{2.51}$$

We can show that, in the steady state (where w grows at the rate x and $r = \rho + \theta x$), the relation $w = \mu\tilde{w}$ holds. Therefore, relative asset positions stay constant in the steady state. Outside of the steady state, equation (2.51) implies that the relative asset position does not change over time for a household whose relative labor productivity, π_j, is as high as its relative asset position, a_j/a. For other households, the behavior depends on the sign of $w - \mu\tilde{w}$. Imagine that $w > \mu\tilde{w}$. Roughly speaking, this condition says that the propensity to save out of (permanent) wage income is positive. In this case, equation (2.51) implies that a_j/a would rise or fall over time depending on whether relative labor productivity exceeded or fell short of the relative asset position—$\pi_j >$(or $<$) a_j/a. Thus a convergence pattern would hold, whereby relative assets moved toward relative productivity. However, the opposite pattern applies if $w < \mu\tilde{w}$. Outside of the steady state, the sign of $w - \mu\tilde{w}$ depends on the relation of interest rates to growth rates of wages and is ambiguous. Hence, the model does not have clear predictions about the way in which a_j/a will move along the transition.

Caselli and Ventura (2000) also allowed for a form of heterogeneity in household preferences. They assumed that preferences involved the felicity function $u(c + \beta_j g)$, where they interpret g as a publicly provided service. The parameter $\beta_j > 0$ indicates the value that household j attaches to the public service. The variable g could also represent the services that households get freely from the environment, for example, from staring at the sky. The main result from this extension is that the aggregation of individual behavior still corresponds to a representative-agent model, in the sense that the economy-wide average variables, a and c, evolve as they would with a single agent who had average values of initial assets, labor productivity, and preferences. In this sense, the results from the Ramsey model are robust to this extension to admit heterogeneous preferences.

2.7 Nonconstant Time-Preference Rates

Many of the basic frameworks in macroeconomics, including the neoclassical growth model that we have been analyzing, rely on the assumption that households have a constant rate of time preference, ρ. However, the rationale for this assumption is unclear.[29] Perhaps it is unclear because the reason for individuals to have positive time preference is itself unclear.

Ramsey (1928, p. 543) preferred to use a zero rate of time preference. He justified this approach in a normative context by saying "we do not discount later enjoyments in comparison with earlier ones, a practice which is ethically indefensible." Similarly, Fisher (1930, chapter 4) argued that time preference—or impatience, as he preferred to call it—reflects mainly a person's lack of foresight and self-control. One reason that economists have not embraced a zero rate of time preference is that it causes difficulties for the long-run equilibrium—in particular, the transversality condition in the model that we have analyzed requires the inequality $\rho > x \cdot (1 - \theta) + n$, which is positive if $\theta < 1 + (n/x)$. Thus most analyses have assumed that the rate of time preference is positive but constant.

29. See Koopmans (1960) and Fishburn and Rubinstein (1982) for axiomatic derivations of a constant rate of time preference.

As has been known since Strotz (1956) and the elaborations of Pollak (1968) and Goldman (1980)—and understood much earlier by Ramsey (1928)[30]—nonconstancy of the rate of time preference can create a time-consistency problem. This problem arises because the relative valuation of utility flows at different dates changes as the planning date evolves. In this context, committed choices of consumption typically differ from those chosen sequentially, taking account of the way that future consumption will be determined. Therefore, the commitment technology matters for the outcomes.

Laibson (1997a, 1997b), motivated partly by introspection and partly by experimental findings, has made compelling observations about ways in which rates of time preference vary.[31] He argues that individuals are highly impatient about consuming between today and tomorrow but are much more patient about choices advanced further in the future, for example, between 365 and 366 days from now. Hence, rates of time preference would be very high in the short run but much lower in the long run, as viewed from today's perspective. Given these insights and evidence, it is important to know whether economists can continue to rely on the standard version of the neoclassical growth model—the model analyzed in this chapter—as their workhorse framework for dynamic macroeconomics.

To assess this issue, we follow the treatment in Barro (1999) and modify the utility function from equation (2.1) to

$$U(\tau) = \int_{\tau}^{\infty} u[c(t)] \cdot e^{-[\rho\cdot(t-\tau)+\phi(t-\tau)]}\, dt \tag{2.52}$$

where τ now represents the current date and $\phi(t-\tau)$ is a function that brings in the aspects of time preference that cannot be described by the standard exponential factor, $e^{-\rho\cdot(t-\tau)}$. For convenience, we begin with a case of zero population growth, $n = 0$, so that the term $e^{n\cdot(t-\tau)}$ does not appear in equation (2.52). We assume that the felicity function takes the usual form given in equation (2.10):

$$u(c) = \frac{c^{(1-\theta)} - 1}{(1-\theta)}$$

30. In the part of his analysis that allows for time preference, Ramsey (1928, p. 439) says, "In assuming the rate of discount constant, I [mean that] the present value of an enjoyment at any future date is to be obtained by discounting it at the rate ρ. . . . This is the only assumption we can make, without contradicting our fundamental hypothesis that successive generations are activated by the same system of preferences. For, if we had a varying rate of discount—say a higher one for the first fifty years—our preference for enjoyments in 2000 A.D. over those in 2050 A.D. would be calculated at the lower rate, but that of the people alive in 2000 A.D. would be at the higher."

31. For discussions of the experimental evidence, see Thaler (1981), Ainslie (1992), and Loewenstein and Prelec (1992).

The new time-preference term, $\phi(t-\tau)$, is assumed, as in the case of the conventional time-preference factor, to depend only on the distance in time, $t-\tau$.[32] We can normalize to have $\phi(0)=0$. We also assume that the function $\phi(\cdot)$ is continuous and twice differentiable. The expression $\rho+\phi'(v)$ gives the instantaneous rate of time preference at the time distance $v=t-\tau\geq 0$. The assumed properties, which follow Laibson (1997a), are $\phi'(v)\geq 0, \phi''(v)\leq 0$, and $\phi'(v)$ approaches zero as v tends to infinity. These properties imply that the rate of time preference, given by $\rho+\phi'(t-\tau)$, is high in the near term but roughly constant at the lower value ρ in the distant future. Consumers with these preferences are impatient about consuming right now, but they need not be shortsighted in the sense of failing to take account of long-term consequences. The analysis assumes no decision-making failures of this sort.

Except for the modification of the time-preference rate, the model is the same as before, including the specification of the production function and the behavior of firms. For convenience, we begin with the case of zero technological change, $x=0$.

2.7.1 Results under Commitment

The first-order optimization conditions for the household's path of consumption, $c(t)$, would be straightforward if the full path of current and future consumption could be chosen in a committed manner at the present time, τ. In particular, the formula for the growth rate of consumption would be modified from equation (2.11) to

$$\dot{c}/c=(1/\theta)\cdot[r(t)-\rho-\phi'(t-\tau)] \tag{2.53}$$

for $t>\tau$. The new element is the addition of the term $\phi'(t-\tau)$ to ρ. Equation (2.53) can be viewed as coming from usual perturbation arguments, whereby consumption is lowered at some point in time and raised at another point in time—perhaps the next instant in time—with all other values of consumption held constant.

Given the assumed properties for $\phi(\cdot)$, $\rho+\phi'(t-\tau)$ would start at a high value and then decline toward ρ as $t-\tau$ tended toward infinity. Thus the steady-state rate of time preference would be ρ, and the steady state of the model would coincide with the analysis from before. The new results would involve the transition, during which time-preference rates were greater than ρ but falling over time.

One problem with this solution is that the current time, τ, is arbitrary, and, in the typical situation, the potential to commit did not suddenly arise at this time. Rather, if perpetual commitments on consumption were feasible, these commitments would likely have existed

32. The utility expression can be extended without affecting the basic results to include the chronological date, t, and a household's age and other life-cycle characteristics.

in the past, perhaps in the infinite past. In this last situation, current and all future values of consumption would have been determined earlier, and τ would be effectively minus infinity, so that $\phi'(t-\tau)$ would be zero for all $t \geq 0$. Hence, the rate of time preference would equal ρ for all $t \geq 0$, and the standard Ramsey results would apply throughout, not just in the steady state.

The more basic problem is that commitment on future choices of $c(t)$ is problematic. The next section therefore works out the solution in the absence of any commitment technology for future consumption. In this setting, the household can determine at time τ only the instantaneous flow of consumption, $c(\tau)$.

2.7.2 Results without Commitment under Log Utility

The first-order condition in equation (2.53) will not generally hold without commitment, because it is infeasible for the household to carry out the perturbations that underlie the condition. Specifically, the household cannot commit to lowering $c(\tau)$ at time τ and increasing $c(t)$ at some future date, while holding fixed consumption at all other dates. Instead, the household has to figure out how its setting of $c(\tau)$ at time τ will alter its stock of assets and how this change in assets will influence the choices of consumption at later dates.

The full solution without commitment is worked out first for log utility, where $\theta = 1$. The steady-state results for general θ are discussed in a later section. Transitional results for general θ are more complicated, but some results are sketched later.

Think of choosing $c(t)$ at time τ as the constant flow $c(\tau)$ over the short discrete interval $[\tau, \tau+\epsilon]$. The length of the interval, ϵ, will eventually approach zero and thereby generate results for continuous time. The full integral of utility flows from equation (2.52) can be broken up into two pieces:

$$U(\tau) = \int_{\tau}^{\tau+\epsilon} \log[c(t)] \cdot e^{-[\rho\cdot(t-\tau)+\phi(t-\tau)]}\, dt + \int_{\tau+\varepsilon}^{\infty} \log[c(t)] \cdot e^{-[\rho\cdot(t-\tau)+\phi(t-\tau)]}\, dt$$
$$\approx \epsilon \cdot \log[c(\tau)] + \int_{\tau+\epsilon}^{\infty} \log[c(t)] \cdot e^{-[\rho\cdot(t-\tau)+\phi(t-\tau)]}\, dt \tag{2.54}$$

where the approximation comes from taking $e^{-[\rho\cdot(t-\tau)+\phi(t-\tau)]}$ as equal to unity over the interval $[\tau, \tau+\epsilon]$. This approximation will become exact in the equilibrium as ϵ tends to zero. Note that log utility has been assumed.[33]

The consumer can pick $c(\tau)$ and thereby the choice of saving at time τ. This selection influences $c(t)$ for $t \geq \tau+\epsilon$ by affecting the stock of assets, $k(\tau+\epsilon)$, available at time

33. Pollak (1968, section 2) works out results under log utility with a finite horizon and a zero interest rate.

$\tau + \epsilon$. (Solely for convenience, we already assume the equality between per capita assets, $a[t]$, and the per capita capital stock, $k[t]$.) To determine the optimal $c(\tau)$, the household has to know, first, the relation between $c(\tau)$ and $k(\tau + \epsilon)$ and, second, the relation between $k(\tau + \epsilon)$ and the choices of $c(t)$ for $t \geq \tau + \epsilon$.

The first part of the problem is straightforward. The household's budget constraint is

$$\dot{k}(t) = r(t) \cdot k(t) + w(t) - c(t) \tag{2.55}$$

For a given starting stock of assets, $k(\tau)$, the stock at time $\tau + \epsilon$ is determined by

$$k(\tau + \epsilon) \approx k(\tau) \cdot [1 + \epsilon \cdot r(\tau)] + \epsilon \cdot w(\tau) - \epsilon \cdot c(\tau) \tag{2.56}$$

The approximation comes from neglecting compounding over the interval $(\tau, \tau + \epsilon)$—that is, from ignoring terms of the order of ϵ^2—and from treating the variables $r(t)$ and $w(t)$ as constants over this interval. These assumptions will be satisfactory in the equilibrium when ϵ approaches zero. The important result from equation (2.56) is that

$$d[k(\tau + \epsilon)]/d[c(\tau)] \approx -\epsilon \tag{2.57}$$

Hence, more consumption today means less assets at the next moment in time.

The difficult calculation involves the link between $k(\tau + \epsilon)$ and $c(t)$ for $t \geq \tau + \epsilon$, that is, the propensities to consume out of assets. In the standard model with log utility, we know from equations (2.15) and (2.16) that—because of the cancellation of income and substitution effects related to the path of interest rates—consumption is a constant fraction of wealth:

$$c(t) = \rho \cdot [k(t) + \tilde{w}(t)]$$

where $\tilde{w}(t)$ is the present value of wages. Given this background, it is reasonable to conjecture that the income and substitution effects associated with interest rates would still cancel under log utility, even though the rate of time preference is variable and commitment is absent. However, the constant of proportionality, denoted by λ, need not equal ρ. Thus, the conjecture—which turns out to be correct—is that consumption is given by

$$c(t) = \lambda \cdot [k(t) + \tilde{w}(t)] \tag{2.58}$$

for $t \geq \tau + \epsilon$ for some constant $\lambda > 0$.[34]

34. Phelps and Pollak (1968, section 4) use an analogous conjecture to work out a Cournot–Nash equilibrium for their problem. They assume isoelastic utility and a linear technology, so that the rate of return is constant. The last property is critical, because consumption is not a constant fraction of wealth (except when $\theta = 1$) if the rate of return varies over time. The linear technology also eliminates any transitional dynamics, so that the economy is always in a position of steady-state growth.

Under the assumed conjecture, it can be verified that $c(t)$ grows at the rate $r(t) - \lambda$ for $t \geq \tau + \epsilon$. Hence, for any $t \geq \tau + \epsilon$, consumption is determined from

$$\log[c(t)] = \log[c(\tau+\epsilon)] + \int_{\tau+\epsilon}^{t} r(v)\, dv - \lambda \cdot (t - \tau - \epsilon)$$

The expression for utility from equation (2.54) can therefore be written as

$$U(\tau) \approx \epsilon \cdot \log[c(\tau)] + \log[c(\tau+\epsilon)] \cdot \int_{\tau+\epsilon}^{\infty} e^{-[\rho\cdot(t-\tau)+\phi(t-\tau)]}\, dt$$
$$+ \text{terms that are independent of } c(t) \text{ path} \tag{2.59}$$

Define the integral

$$\Omega(\epsilon) \equiv \int_{\epsilon}^{\infty} e^{-[\rho v+\phi(v)]}\, dv \tag{2.60}$$

The marginal effect of $c(\tau)$ on $U(\tau)$ can then be calculated as

$$\frac{d[U(\tau)]}{d[c(\tau)]} \approx \frac{\epsilon}{c(\tau)} + \frac{\Omega(\epsilon)}{c(\tau+\epsilon)} \cdot \frac{d[c(\tau+\epsilon)]}{d[k(\tau+\epsilon)]} \cdot \frac{d[k(\tau+\epsilon)]}{dc(\tau)}$$

The final derivative equals $-\epsilon$, from equation (2.57), and the next-to-last derivative equals λ, according to the conjectured solution in equation (2.58). Therefore, setting $d[U(\tau)]/d[c(\tau)]$ to zero implies

$$c(\tau) = \frac{c(\tau+\epsilon)}{\lambda \cdot \Omega(\epsilon)}$$

If the conjectured solution is correct, $c(\tau+\epsilon)$ must approach $c(\tau)$ as ϵ tends to zero. Otherwise, $c(t)$ would exhibit jumps at all points in time, and the conjectured answer would be wrong. The unique value of λ that delivers this correspondence follows immediately as

$$\lambda = 1/\Omega = \frac{1}{\int_0^{\infty} e^{-[\rho v+\phi(v)]}\, dv} \tag{2.61}$$

where we use the notation $\Omega \equiv \Omega(0)$.

To summarize, the solution for the household's consumption problem under log utility is that $c(t)$ be set as the fraction λ of wealth at each date, where λ is the constant shown in equation (2.61). The solution is time consistent because, if $c(t)$ is chosen in this manner at all future dates, it will be optimal for consumption to be set this way at the current date.[35]

Inspection of equation (2.61) reveals that $\lambda = \rho$ in the standard Ramsey case in which $\phi(v) = 0$ for all v. To assess the general implications of $\phi(v)$ for λ, it is convenient to rewrite equation (2.62) as

$$\lambda = \frac{\int_0^\infty e^{-[\rho v + \phi(v)]} \cdot [\rho + \phi'(v)]\, dv}{\int_0^\infty e^{-[\rho v + \phi(v)]}\, dv} \tag{2.62}$$

Since the numerator of equation (2.62) equals unity,[36] this result corresponds to equation (2.61).

The form of equation (2.62) is useful because it shows that λ is a time-invariant weighted average of the instantaneous rates of time preference, $\rho + \phi'(v)$. Since $\phi'(v) \geq 0, \phi''(v) \leq 0$, and $\phi'(v) \to 0$ as $v \to \infty$, it follows that

$$\rho \leq \lambda \leq \rho + \phi'(0) \tag{2.63}$$

That is, λ is intermediate between the long-run rate of time preference, ρ, and the short-run, instantaneous rate, $\rho + \phi'(0)$.

The determination of the effective rate of time preference can be quantified by specifying the form of $\phi(v)$. Laibson (1997a) proposes a "quasi-hyperbola" in discrete time, whereby $\phi(v) = 0$ in the current period and $e^{-\phi(v)} = \beta$ in each subsequent period, where $0 < \beta \leq 1$. (Phelps and Pollak, 1968, also use this form.) In this specification, the discount factor between today and tomorrow includes the factor $\beta \leq 1$. This factor does not enter between any two adjacent future periods. Laibson argues that β would be substantially less than one on an annual basis, perhaps between one-half and two-thirds.

This quasi-hyperbolic case can be applied to a continuous-time setting by specifying

$$\phi(v) = 0 \text{ for } 0 \leq v \leq V, \qquad e^{-\phi(v)} = \beta \text{ for } v > V \tag{2.64}$$

35. This approach derives equation (2.61) as a Cournot–Nash equilibrium but does not show that the equilibrium is unique. Uniqueness is easy to demonstrate in the associated discrete-time model with a finite horizon, as considered by Laibson (1996). In the final period, the household consumes all of its assets, and the unique solution for each earlier period can be found by working backward sequentially from the end point. This result holds as long as $u(c)$ is concave, not just for isoelastic utility. The uniqueness result also holds if the length of a period approaches zero (to get continuous time) and if the length of the horizon becomes arbitrarily large. However, Laibson (1994) uses an explicitly game-theoretic approach to demonstrate the possibility of nonuniqueness of equilibrium in the infinite-horizon case. The existence of multiple equilibria depends on punishments that sanction past departures of consumption choices from designated values, and these kinds of equilibria unravel if the horizon is finite. Our analysis of the infinite-horizon case does not consider these kinds of equilibria.

36. Use the change of variable $z = e^{-[\rho v + \phi(v)]}$.

for some $V > 0$, where $0 < \beta \leq 1$. [In this specification, $\phi'(v)$ is infinite at $v = V$ and equals zero otherwise.] Laibson's suggestion is that V is small, so that the condition $\rho V \ll 1$ would hold.

Substitution from equation (2.64) into the definition of Ω in equation (2.60) leads (when $\epsilon = 0$) to

$$\Omega = (1/\rho) \cdot [1 - (1 - \beta) \cdot e^{-\rho V}]$$

As V approaches infinity, Ω goes to $1/\rho$, which corresponds to the Ramsey case. The condition $\rho V \ll 1$ implies that the expression for Ω simplifies, as an approximation, to β/ρ, so that

$$\lambda \approx \rho/\beta \tag{2.65}$$

If β is between one-half and two-thirds, λ is between 1.5ρ and 2ρ. Hence, if ρ is 0.02 per year, the heavy near-term discounting of future utility converts the Ramsey model into one in which the effective rate of time preference, λ, is 0.03–0.04 per year.

The specification in equation (2.64) yields simple closed-form results, but the functional form implies an odd discrete jump in $e^{-\phi(v)}$ at the time V in the future. More generally, the notion from the literature on short-term impatience is that $\rho + \phi'(v)$ is high when v is small and declines, say toward ρ, as v becomes large. A simple functional form that captures this property in a smooth fashion is

$$\phi'(v) = be^{-\gamma v} \tag{2.66}$$

where $b = \phi'(0) \geq 0$ and $\gamma > 0$. The parameter γ determines the constant rate at which $\phi'(v)$ declines from $\phi'(0)$ to zero.

Integration of the expression in equation (2.66), together with the boundary condition $\phi(0) = 0$, leads to an expression for $\phi(v)$:[37]

$$\phi(v) = (b/\gamma) \cdot (1 - e^{-\gamma v}) \tag{2.67}$$

This result can be substituted into the formula in equation (2.60) to get an expression for Ω:

$$\Omega = e^{-(b/\gamma)} \cdot \int_0^\infty e^{[-\rho v + (b/\gamma) \cdot e^{-\gamma v}]} \, dv$$

The integral cannot be solved in closed form but can be evaluated numerically if values are specified for the parameters ρ, b, and γ.

37. The expression in equation (2.67) is similar to the "generalized hyperbola" proposed by Loewenstein and Prelec (1992, p. 580). Their expression can be written as $\phi(v) = (b/\gamma) \cdot \log(1 + \gamma v)$.

To accord with Laibson's (1997a) observations, the parameter $b = \phi'(0)$ must be around 0.50 per year, and the parameter γ must be at least 0.50 per year, so that $\phi'(v)$ gets close to zero a few years in the future. With $\rho = 0.02$, $b = 0.50$, and $\gamma = 0.50$, Ω turns out to be 19.3, so that $\lambda = 1/\Omega = 0.052$. If $b = 0.25$ and the other parameters are the same, $\Omega = 31.0$ and $\lambda = 0.032$. Thus, the more appealing functional form in equation (2.67) has implications that are similar to those of equation (2.64).

The introduction of the $\phi(\cdot)$ term in the utility function of equation (2.52) and the consequent shift to a time-inconsistent setting amount, under log utility, to an increase in the rate of time preference above ρ. Since the effective rate of time preference, λ, is constant, the dynamics and steady state of the model take exactly the same form as in the standard Ramsey framework that we analyzed before. The higher rate of time preference corresponds to a higher steady-state interest rate,

$$r^* = \lambda \tag{2.68}$$

and, thereby, to a lower steady-state capital intensity, k^*, which is determined from the condition

$$f'(k^*) = \lambda + \delta$$

Since the effective rate of time preference, λ, is constant, the model with log utility and no commitment is observationally equivalent to the conventional neoclassical growth model. That is, the equilibrium coincides with that in the standard model for a suitable choice of ρ. Since the parameter ρ cannot be observed directly, there is a problem in inferring from data whether the instantaneous rate of time preference includes the nonconstant term, $\phi'(v)$.

2.7.3 Population Growth and Technological Progress

It is straightforward to incorporate population growth in the manner of equation (2.1). The solution under log utility is similar to that from before, except that the integral Ω is now defined by

$$\Omega \equiv \int_0^{\infty} e^{-[(\rho-n)\cdot v+\phi(v)]}\, dv \tag{2.69}$$

The relation between the propensity to consume out of wealth, λ, and the modified Ω term is given by

$$\lambda = n + (1/\Omega) \tag{2.70}$$

and the steady-state interest rate is again $r^* = \lambda$. We leave the derivations of these results as exercises.

In the Ramsey case, where $\phi(v) = 0$ for all v, $\Omega = 1/(\rho - n)$ in equation (2.69) and $\lambda = \rho$ in equation (2.70). For Laibson's quasi-hyperbolic preferences in equation (2.64), the result is

$$\Omega \approx \beta/(\rho - n), \qquad \lambda \approx (\rho/\beta) - n \cdot (1 - \beta)/\beta \tag{2.71}$$

If $0 < \beta < 1$, an increase in n lowers λ and, therefore, reduces the steady-state interest rate, $r^* = \lambda$.

It is also straightforward to introduce exogenous, labor-augmenting technological progress at the rate $x \geq 0$. The solution for λ is still that shown in equations (2.69) and (2.70). However, since consumption per person grows in the steady state at the rate x, the condition for the steady-state interest rate is

$$r^* = \lambda + x$$

Hence, as is usual with log utility, r^* responds one-to-one to the rate of technological progress, x.

2.7.4 Results under Isoelastic Utility

In the standard analysis, where $\phi(t - \tau) = 0$ for all t, consumption is not a constant fraction of wealth unless $\theta = 1$. However, we know, for any value of θ, that the first-order condition for consumption growth at time τ is given from equation (2.11) by

$$\frac{\dot{c}}{c}(\tau) = (1/\theta) \cdot [r(\tau) - \rho] \tag{2.72}$$

A reasonable conjecture is that the form of equation (2.72) would still hold when $\phi(t - \tau) \neq 0$ but that the constant ρ would be replaced by some other constant that represented the effective rate of time preference. This conjecture is incorrect. The reason is that the effective rate of time preference at time τ involves an interaction of the path of the future values of $\phi'(t - \tau)$ with future interest rates and turns out not to be constant when interest rates are changing except when $\theta = 1$.

Although the transitional dynamics is complicated, it is straightforward to work out the characteristics of the steady state. The key point is that, in a steady state, an increase in household assets would be used to raise consumption uniformly in future periods. This property makes it easy to compute propensities to consume for future periods with respect to current assets and, therefore, makes it easy to find the first-order optimization condition for current consumption. Only the results are presented here.

In the steady state, the interest rate is given by

$$r^* = x + n + 1/\Omega \tag{2.73}$$

where the integral Ω is now defined by

$$\Omega \equiv \int_0^{\infty} e^{-\{[\rho - x\cdot(1-\theta)-n]\cdot v + \phi(v)\}}\, dv \tag{2.74}$$

Thus, if $\phi(v) = 0$, we get the standard result

$$r^* = \rho + \theta x$$

For the case of Laibson's quasi-hyperbolic utility function in equation (2.64), the result turns out to be

$$r^* \approx \frac{\rho}{\beta} - n \cdot \frac{(1-\beta)}{\beta} + x \cdot \frac{(\beta + \theta - 1)}{\beta} \tag{2.75}$$

where recall that $0 < \beta < 1$. Thus, for the case considered before of log utility ($\theta = 1$), the effect of x on r^* is one-to-one. More generally, the effect of x on r^* is more or less than one-to-one depending on whether θ is greater or less than 1.

For the transitional dynamics, Barro (1999) shows that consumption growth at any date τ satisfies the condition

$$\frac{\dot{c}}{c}(\tau) = (1/\theta) \cdot [r(\tau) - \lambda(\tau)] \tag{2.76}$$

The term $\lambda(\tau)$ is the effective rate of time preference and is given by

$$\lambda(\tau) = \frac{\int_\tau^\infty \omega(t, \tau) \cdot [\rho + \phi'(t-\tau)]\, dt}{\int_\tau^\infty \omega(t, \tau)\, dt} \tag{2.77}$$

where $\omega(t, \tau) > 0$. Thus, $\lambda(\tau)$ is again a weighted average of future instantaneous rates of time preference, $\rho + \phi'(t - \tau)$. The difference from equation (2.62) is that the weighting factor, $\omega(t, \tau)$, is time varying unless $\theta = 1$.

Barro (1999) shows that, if $\theta > 1$, $\omega(t, \tau)$ declines with the average of interest rates between dates τ and t. If the economy begins with a capital intensity below its steady-state value, $r(\tau)$ starts high and then falls toward its steady-state value. The weights $\omega(t, \tau)$ are then particularly low for dates t far in the future. Since these dates are also the ones with relatively low values of $\rho + \phi'(t - \tau)$, $\lambda(\tau)$ is high initially. However, as interest rates fall, the weights, $\omega(t, \tau)$, become more even, and $\lambda(\tau)$ declines. This descending path of $\lambda(\tau)$ means that households effectively become more patient over time. However, the effects are all reversed if $\theta < 1$. The case $\theta = 1$, which we worked out before, is the intermediate one in which the weights stay constant during the transition. Hence, in this case, the effective rate of time preference does not change during the transition.

2.7.5 The Degree of Commitment

The analysis thus far considered a case of full commitment, as in equation (2.53), and ones of zero commitment, as in equation (2.76). Barro (1999) also considers intermediate cases in which commitment is possible over an interval of length T, where $0 \leq T \leq \infty$. Increases in the extent of commitment—that is, higher T—lead in the long run to a lower effective rate of time preference and, hence, to lower interest rates and higher capital intensity. However, changes in T also imply transitional effects—initially an increase in T tends to make households *less* patient because they suddenly get the ability to constrain their "future selves" to save more. Thus the analysis implies that a rise in T initially lowers the saving rate but tends, in the longer run, to raise the willingness to save.

If the parameter T can be identified with observable variables—such as the nature of legal and financial institutions or cultural characteristics that influence the extent of individual discipline—the new theoretical results might eventually have empirical application. In fact, from an empirical standpoint, the main new insights from the extended model concern the connection between the degree of commitment and variables such as interest rates and saving rates. For a given degree of commitment, the main result is that a nonconstant rate of time preference leaves intact the main implications of the neoclassical growth model.

2.8 Appendix 2A: Log-Linearization of the Ramsey Model

The system of differential equations that characterizes the Ramsey model is given from equations (2.24) and (2.25) by

$$\dot{\hat{k}} = f(\hat{k}) - \hat{c} - (x + n + \delta) \cdot \hat{k}$$

$$\dot{\hat{c}}/\hat{c} = \dot{c}/c - x = (1/\theta) \cdot [f'(\hat{k}) - \delta - \rho - \theta x] \tag{2.78}$$

We now log-linearize this system for the case in which the production function is Cobb–Douglas, $f(\hat{k}) = A \cdot \hat{k}^{\alpha}$.

Start by rewriting the system from equation (2.78) in terms of the logs of $\hat{c}$ and $\hat{k}$:

$$\begin{aligned} d[\log(\hat{k})]/dt &= A \cdot e^{-(1-\alpha)\cdot\log(\hat{k})} - e^{\log(\hat{c}/\hat{k})} - (x + n + \delta) \\ d[\log(\hat{c})]/dt &= (1/\theta) \cdot [\alpha A \cdot e^{-(1-\alpha)\cdot\log(\hat{k})} - (\rho + \theta x + \delta)] \end{aligned} \tag{2.79}$$

In the steady state, where $d[\log(\hat{k})]/dt = d[\log(\hat{c})]/dt = 0$, we have

$$A \cdot e^{-(1-\alpha)\cdot\log(\hat{k}^*)} - e^{\log(\hat{c}^*/\hat{k}^*)} = (x + n + \delta)$$

$$\alpha A \cdot e^{-(1-\alpha)\cdot\log(\hat{k}^*)} = (\rho + \theta x + \delta) \tag{2.80}$$

We take a first-order Taylor expansion of equation (2.79) around the steady-state values determined by equation (2.80):

$$\begin{bmatrix} d[\log(\hat{k})]/dt \\ d[\log(\hat{c})]/dt \end{bmatrix} = \begin{bmatrix} \zeta & x+n+\delta-\dfrac{(\rho+\theta x+\delta)}{\alpha} \\ -(1-\alpha)\cdot\dfrac{(\rho+\theta x+\delta)}{\theta} & 0 \end{bmatrix} \cdot \begin{bmatrix} \log(\hat{k}/\hat{k}^*) \\ \log(\hat{c}/\hat{c}^*) \end{bmatrix} \tag{2.81}$$

where $\zeta \equiv \rho - n - (1-\theta)\cdot x$. The determinant of the characteristic matrix equals

$$-[(\rho+\theta x+\delta)/\alpha - (x+n+\delta)]\cdot(\rho+\theta x+\delta)\cdot(1-\alpha)/\theta$$

Since $\rho + \theta x > x + n$ (from the transversality condition in equation [2.31]) and $\alpha < 1$, the determinant is negative. This condition implies that the two eigenvalues of the system have opposite signs, a result that implies saddle-path stability. (See the discussion in the mathematics appendix at the end of the book.)

To compute the eigenvalues, denoted by ϵ, we use the condition

$$\det\begin{bmatrix} \zeta-\epsilon & x+n+\delta-\dfrac{(\rho+\theta x+\delta)}{\alpha} \\ -(1-\alpha)\cdot\dfrac{(\rho+\theta x+\delta)}{\theta} & -\epsilon \end{bmatrix} = 0 \tag{2.82}$$

This condition corresponds to a quadratic equation in ϵ :

$$\epsilon^2 - \zeta\cdot\epsilon - [(\rho+\theta x+\delta)/\alpha - (x+n+\delta)]\cdot[(\rho+\theta x+\delta)\cdot(1-\alpha)/\theta] = 0 \tag{2.83}$$

This equation has two solutions:

$$2\epsilon = \zeta \pm \left[\zeta^2 + 4\cdot\left(\frac{1-\alpha}{\theta}\right)\cdot(\rho+\theta x+\delta)\cdot[(\rho+\theta x+\delta)/\alpha - (x+n+\delta)]\right]^{1/2} \tag{2.84}$$

where ϵ_1, the root with the positive sign, is positive, and ϵ_2, the root with the negative sign, is negative. Note that ϵ_2 corresponds to $-\beta$ in equation (2.41).

The log-linearized solution for $\log(\hat{k})$ takes the form

$$\log[\hat{k}(t)] = \log(\hat{k}^*) + \psi_1\cdot e^{\epsilon_1 t} + \psi_2\cdot e^{\epsilon_2 t} \tag{2.85}$$

where ψ_1 and ψ_2 are arbitrary constants of integration. Since $\epsilon_1 > 0$, $\psi_1 = 0$ must hold for $\log[\hat{k}(t)]$ to tend asymptotically to $\log(\hat{k}^*)$. ($\psi_1 > 0$ violates the transversality condition, and $\psi_1 < 0$ leads to $\hat{k} \to 0$, which corresponds to cases in which the system hits the vertical axis in figure 2.1.) The other constant, ψ_2, is determined from the initial condition:

$$\psi_2 = \log[\hat{k}(0)] - \log(\hat{k}^*) \tag{2.86}$$

If we substitute $\psi_1 = 0$, the value of ψ_2 from equation (2.86), and $\epsilon_2 = -\beta$ into equation (2.85), we get the time path for $\log[\hat{k}(t)]$:

$$\log[\hat{k}(t)] = (1 - e^{-\beta t}) \cdot \log(\hat{k}^*) + e^{-\beta t} \cdot \log[\hat{k}(0)] \tag{2.87}$$

Since $\log[\hat{y}(t)] = \log(A) + \alpha \cdot \log[\hat{k}(t)]$, the time path for $\log[\hat{y}(t)]$ is given by

$$\log[\hat{y}(t)] = (1 - e^{-\beta t}) \cdot \log(\hat{y}^*) + e^{-\beta t} \cdot \log[\hat{y}(0)] \tag{2.88}$$

which corresponds to equation (2.40).

2.9 Appendix 2B: Irreversible Investment

Suppose that investment is irreversible, so that $\hat{c} \leq f(\hat{k})$ applies. Reconsider in this case the dynamic paths that start with $\hat{k} < \hat{k}^*$ at a position such as $\hat{c}_0^{'}$ in figure 2.1. These paths would eventually hit the production function, $\hat{c} = f(\hat{k})$, after which the constraint from irreversible investment would become binding. Thereafter, the paths would move downward along the production function, so that $\hat{c} = f(\hat{k})$ would apply. Hence, the capital intensity would decline in accordance with $\dot{\hat{k}} = -(x + n + \delta) \cdot \hat{k}$. Therefore, $\hat{k}$ (and $\hat{c}$) would asymptotically approach zero but would not reach zero in finite time. We now argue that such paths cannot be equilibria.

When the constraint $\hat{c} \leq f(\hat{k})$ is binding, so that all output goes to consumption and none to gross investment, the price of capital, denoted by ϕ, can fall below 1. The rate of return to holders of capital then satisfies (see note 11)

$$r = R/\phi - \delta + \dot{\phi}/\phi \tag{2.89}$$

Profit maximization for competitive firms still implies the condition $R = f'(\hat{k})$, which can be substituted into the formula for r.

Consumer optimization entails, as usual,

$$\dot{c}/c = (1/\theta) \cdot (r - \rho)$$

Therefore, substitution for r from equation (2.89) yields the formula for the growth rate of $\hat{c}$:

$$\dot{\hat{c}}/\hat{c} = \left(\frac{1}{\theta\phi}\right) \cdot [f'(\hat{k}) + \dot{\phi} - \phi \cdot (\delta + \rho + \theta x)] \tag{2.90}$$

The condition $\hat{c} = f(\hat{k})$, together with $\dot{\hat{k}} = -(x+n+\delta)\cdot\hat{k}$, implies another condition for the growth rate of $\hat{c}$:

$$\dot{\hat{c}}/\hat{c} = -\alpha(\hat{k})\cdot(x+n+\delta) \tag{2.91}$$

where $\alpha(\hat{k}) \equiv \hat{k}\cdot f'(\hat{k})/f(\hat{k})$ is the capital share of income (which is a constant in the case of a Cobb–Douglas production function). Therefore, equations (2.90) and (2.91) imply a condition for $\dot{\phi}$:

$$\dot{\phi} = -f'(\hat{k}) + \phi\cdot[\delta+\rho+\theta x - \alpha(\hat{k})\cdot\theta\cdot(x+n+\delta)] \tag{2.92}$$

Suppose that the constraint $\hat{c} \leq f(\hat{k})$ first becomes binding at some date T, where $\hat{k}(T) < \hat{k}^*$ applies. At this point, $f'(\hat{k}) - \delta > \rho + \theta x$. Therefore, when $\phi = 1$ (just at time T), equation (2.92) implies that $\dot{\phi} < 0$. Over time, the rise in R and the fall in ϕ tend to raise r in accordance with equation (2.81). Nevertheless, households are satisfied with a negative growth rate of $\hat{c}$ (equation [2.91]) because the rate of capital loss, $\dot{\phi}/\phi$, rises sufficiently in magnitude to maintain a low rate of return, r. However, equation (2.92) implies, as $\hat{k}$ decreases and $f'(\hat{k})$ rises, that $\dot{\phi}$ eventually rises in magnitude toward infinity (regardless of what happens to $\alpha[\hat{k}]$ in the range between 0 and 1). Therefore, ϕ would reach zero in finite time and then become negative. This condition violates free disposal with respect to claims on capital. Hence, paths in which the irreversibility constraint, $\hat{c} \leq f(\hat{k})$, is binding cannot exist in the region where $\hat{k} < \hat{k}^*$.

The constraint $\hat{c} \leq f(\hat{k})$ can be binding in the region where $\hat{k} > \hat{k}^*$. This possibility was noted and discussed by Arrow and Kurz (1970).

2.10 Appendix 2C: Behavior of the Saving Rate

This section provides an algebraic treatment of the transitional behavior of the saving rate. We deal here with the transition in which $\hat{k}$ and $\hat{c}$ are rising over time, and we assume a Cobb–Douglas production function, so that $f(\hat{k}) = A\hat{k}^\alpha$.

The gross saving rate, s, equals $1 - \hat{c}/f(\hat{k})$. In the steady state, $\dot{\hat{k}}$ from equation (2.24) and $\dot{\hat{c}}/\hat{c}$ from equation (2.25) are each equal to 0. If we use these conditions, together with $f(\hat{k})/\hat{k} = f'(\hat{k})/\alpha$, which holds in the Cobb–Douglas case, we find that the steady-state saving rate is

$$s^* = \alpha\cdot(x+n+\delta)/(\rho+\theta x+\delta) \tag{2.93}$$

The transversality condition in equation (2.31) implies $\rho+\theta x > x+n$ and, therefore, $s^* < \alpha$.

Since $s = 1 - \hat{c}/f(\hat{k})$, s moves in the direction opposite to the consumption ratio, $\hat{c}/f(\hat{k})$. Define $z \equiv \hat{c}/f(\hat{k})$ and differentiate the ratio to get

$$\gamma_z \equiv \dot{z}/z = \dot{\hat{c}}/\hat{c} - \frac{f'(\hat{k}) \cdot \dot{\hat{k}}}{f(\hat{k})} = \dot{\hat{c}}/\hat{c} - \alpha \cdot (\dot{\hat{k}}/\hat{k}) \tag{2.94}$$

where the last term on the right follows in the Cobb–Douglas case. Substitution from equations (2.24) and (2.25) into equation (2.94) leads to

$$\gamma_z = f'(\hat{k}) \cdot [z(t) - (\theta - 1)/\theta] + (\delta + \rho + \theta x) \cdot (s^* - 1/\theta) \tag{2.95}$$

where we used the condition $f(\hat{k})/\hat{k} = f'(\hat{k})/\alpha$, which holds in the Cobb–Douglas case.

The behavior of z depends on whether s^* is greater than, equal to, or less than $1/\theta$. Suppose first that $s^* = 1/\theta$. Then $z(t) = (\theta - 1)/\theta$ is consistent with $\gamma_z = 0$ in equation (2.95). In contrast, $z(t) > (\theta - 1)/\theta$ for some t would imply $\gamma_z > 0$ for all t, a result that is inconsistent with z approaching its steady-state value. Similarly, $z(t) < (\theta - 1)/\theta$ can be ruled out because it implies $\gamma_z < 0$ for all t. Therefore, if $s^* = 1/\theta$, z is constant at the value $(\theta - 1)/\theta$, and, hence, the saving rate, s, equals the constant $1/\theta$. By analogous reasoning, we find that $s^* > 1/\theta$ implies $z(t) < (\theta - 1)/\theta$ for all t, whereas $s^* < 1/\theta$ implies $z(t) > (\theta - 1)/\theta$ for all t.

Differentiation of equation (2.95) with respect to time implies

$$\dot{\gamma}_z = f''(\hat{k}) \cdot (\dot{\hat{k}}) \cdot [z(t) - (\theta - 1)/\theta] + f'(\hat{k}) \cdot \gamma_z \cdot z(t) \tag{2.96}$$

Suppose now that $s^* > 1/\theta$, so that $z(t) < (\theta - 1)/\theta$ holds for all t. Then $\gamma_z > 0$ for some t would imply $\dot{\gamma}_z > 0$ in equation (2.96) (because $f''(\hat{k}) < 0$, $f'(\hat{k}) > 0$, and $\dot{\hat{k}} > 0$). Therefore, $\gamma_z > 0$ would apply for all t, a result that is inconsistent with the economy's approaching a steady state. It follows if $s^* > 1/\theta$ that $\gamma_z < 0$, and, hence, $\dot{s} > 0$. By an analogous argument, $\gamma_z > 0$ and $\dot{s} < 0$ must hold if $s^* < 1/\theta$.

The results can be summarized as follows:

$s^* = 1/\theta$ implies $s(t) = 1/\theta$, a constant

$s^* > 1/\theta$ implies $s(t) > 1/\theta$ and $\dot{s}(t) > 0$

$s^* < 1/\theta$ implies $s(t) < 1/\theta$ and $\dot{s}(t) < 0$

These results are consistent with the graphical presentation in figure 2.3.

If we use the formula for s^* from equation (2.93), we find that $s^* \geq 1/\theta$ requires $\theta \geq (\rho + \theta x + \delta)/[\alpha \cdot (x + n + \delta)] > 1/\alpha$. Therefore, if $\theta \leq 1/\alpha$, the parameters must be in the range in which $\dot{s} < 0$ applies throughout. In other words, if $\theta \leq 1/\alpha$, the intertemporal-substitution effect is strong enough to ensure that the saving rate falls during the transition.

However, for our preferred value of α in the neighborhood of 0.75, this inequality requires $\theta \leq 1.33$ and is unlikely to hold.

We can analyze the behavior of the consumption/capital ratio, $\hat{c}/\hat{k}$, in a similar way. The results are as follows:

$\theta = \alpha$ implies $\hat{c}/\hat{k} = (\delta + \rho)/\theta - (\delta + n)$, a constant

$\theta < \alpha$ implies $\hat{c}/\hat{k} < (\delta + \rho)/\theta - (\delta + n)$ and $\hat{c}/\hat{k}$ rising over time

$\theta > \alpha$ implies $\hat{c}/\hat{k} > (\delta + \rho)/\theta - (\delta + n)$ and $\hat{c}/\hat{k}$ falling over time

2.11 Appendix 2D: Proof That $\gamma_{\hat{k}}$ Declines Monotonically If the Economy Starts from $\hat{k}(0) < \hat{k}^*$

We need first to prove the following: $\hat{c}(0)$ declines if $r(v)$ increases over some interval for any $v \geq 0$.[38] Equations (2.15) and (2.16) imply

$$\hat{c}(0) = \frac{\hat{k}(0) + \int_0^\infty \hat{w}(t)e^{-[\bar{r}(t)-n-x]t}\, dt}{\int_0^\infty e^{[\bar{r}(t)\cdot(1-\theta)/\theta-\rho/\theta+n]t}\, dt} \tag{2.97}$$

where $\bar{r}(t)$ is the average interest rate between times 0 and t, as defined in equation (2.13). Higher values of $r(v)$ for any $0 \leq v \leq t$ raise $\bar{r}(t)$ and thereby reduce the numerator in equation (2.97). Higher values of $r(v)$ raise the denominator if $\theta \leq 1$; therefore, the result follows at once if $\theta \leq 1$. Assume now that $\theta > 1$, so that the denominator decreases with an increase in $r(v)$. We know that $r(v)\cdot(1-\theta)/\theta - \rho/\theta + n < 0$ if $\theta > 1$ because $r(v)$ exceeds $\rho + \theta x$, the steady-state interest rate, which exceeds $x + n$ from the transversality condition. Therefore, the denominator in equation (2.97) becomes proportionately more sensitive to $r(v)$ (in the negative direction) the larger the value of θ. Accordingly, if we prove the result for $\theta \to \infty$, the result holds for all $\theta > 0$. Using $\theta \to \infty$, equation (2.97) simplifies to

$$\hat{c}(0) = \frac{\hat{k}(0) + \int_0^\infty \hat{w}(t)e^{-[\bar{r}(t)-x-n]t}\, dt}{\int_0^\infty e^{-[\bar{r}(t)-n]t}\, dt} \tag{2.98}$$

Equation (2.98) can be rewritten as

$$\hat{c}(0) = \frac{\int_0^\infty \psi(t)e^{-[\bar{r}(t)-n-x]t}\, dt}{\int_0^\infty \phi(t)e^{-[\bar{r}(t)-n-x]t}\, dt} \tag{2.99}$$

38. We are grateful to Olivier Blanchard for his help with this part of the proof.

where $\psi(t) = \hat{k}(0) \cdot [r(t) - n - x] + \hat{w}(t)$ and $\phi(t) = e^{-xt}$. The result $\dot{\phi} < 0$ follows immediately, and $\dot{\psi} > 0$ can be shown using the conditions $r(t) = f'[\hat{k}(t)] - \delta$, $\hat{w}(t) = f[\hat{k}(t)] - \hat{k}(t) \cdot f'[\hat{k}(t)]$, $\hat{k}(t) > \hat{k}(0)$, and $\dot{\hat{k}} > 0$. Therefore, an increase in $r(v)$ for $0 \le v \le t$, which raises $\bar{r}(t)$, has a proportionately larger negative effect on the numerator of equation (2.99) than on the denominator. It follows that the net effect of an increase in $r(v)$ on $\hat{c}(0)$ is negative, the result that we need.

We can use this result to get a lower bound for $\hat{c}(0)$. Since $r(0) > \bar{r}(t)$, if we substitute $r(0)$ for $\bar{r}(t)$ and $\hat{w}(0)$ for $\hat{w}(t)$ in equation (2.97), then $\hat{c}(0)$ must go down. Therefore,[39]

$$\hat{c}(0)/\hat{k}(0) > [r(0) \cdot (1-\theta)/\theta + \rho/\theta - n] \cdot \left[1 + \frac{\hat{w}(0)}{\hat{k} \cdot [r(0) - n - x]}\right] \tag{2.100}$$

We shall use this inequality later.

The growth rate of $\hat{k}$ is given from equation (2.24) as

$$\gamma_{\hat{k}} = f(\hat{k})/\hat{k} - \hat{c}/\hat{k} - (x + n + \delta) \tag{2.101}$$

where we now omit the time subscripts. Differentiation of equation (2.101) with respect to time yields

$$\dot{\gamma}_{\hat{k}} = -(\hat{w}/\hat{k}) \cdot \gamma_{\hat{k}} - d(\hat{c}/\hat{k})/dt$$

where we used the condition $\hat{w} = f(\hat{k}) - \hat{k} \cdot f'(\hat{k})$. We want to show that $\dot{\gamma}_{\hat{k}} < 0$ holds in the transition during which $\hat{k}$ and $\hat{c}$ are rising. The formulas for $\dot{\hat{c}}/\hat{c}$ in equation (2.25) and $\dot{\hat{k}}$ in equation (2.24) can be used to get

$$\dot{\gamma}_{\hat{k}} = -(\hat{w}/\hat{k}) \cdot \gamma_{\hat{k}} + (\hat{c}/\hat{k}) \cdot [\hat{w}/\hat{k} + [f'(\hat{k}) - \delta] \cdot (\theta - 1)/\theta + \rho/\theta - n - \hat{c}/\hat{k}] \tag{2.102}$$

Hence, if $\hat{c}/\hat{k} \ge \hat{w}/\hat{k} + [f'(\hat{k}) - \delta] \cdot (\theta - 1)/\theta + \rho/\theta - n$, then $\dot{\gamma}_{\hat{k}} < 0$ follows from $\gamma_{\hat{k}} > 0$, Q.E.D. Accordingly, we now assume

$$\hat{c}/\hat{k} < \hat{w}/\hat{k} + [f'(\hat{k}) - \delta] \cdot (\theta - 1)/\theta + \rho/\theta - n \tag{2.103}$$

If we replace $\hat{c}/\hat{k}$ to the left of the brackets in equation (2.102) by the right-hand side of the inequality in equation (2.103), use the formula for $\gamma_{\hat{k}}$ from equation (2.101), and replace $f(\hat{k})/\hat{k}$ by $\hat{w}/\hat{k} + f'(\hat{k})$, then we eventually get

$$\begin{aligned} \dot{\gamma}_{\hat{k}} < {} & -(\hat{w}/\hat{k}) \cdot [f'(\hat{k}) - \delta - \rho - \theta x]/\theta + [\rho/\theta - n + [f'(\hat{k}) - \delta] \cdot (\theta - 1)/\theta]^2 \\ & + [\rho/\theta - n + [f'(\hat{k}) - \delta] \cdot (\theta - 1)/\theta] \cdot (\hat{w} - \hat{c})/\hat{k} \end{aligned} \tag{2.104}$$

39. The result follows from integration of the right-hand side of equation (2.97) if $[r(0) \cdot (1-\theta)/\theta + \rho/\theta - n] > 0$. If this expression is nonpositive, the inequality in equation (2.100) holds trivially.

If $\rho/\theta - n + [f'(\hat{k}) - \delta] \cdot (\theta - 1)/\theta \le 0$, we can use the inequality in equation (2.103) to show $\dot{\gamma}_{\hat{k}} < 0$, Q.E.D. Therefore, we now assume

$$\rho/\theta - n + [f'(\hat{k}) - \delta] \cdot (\theta - 1)/\theta > 0 \tag{2.105}$$

Given the inequality in equation (2.105), we can use the lower bound for $\hat{c}/\hat{k}$ from equation (2.100) in equation (2.104) to get, after some manipulation,

$$\dot{\gamma}_{\hat{k}} < -\frac{(\hat{w}/\hat{k}) \cdot [f'(\hat{k}) - \delta - \rho - \theta x]^2}{[f'(\hat{k}) - \delta - n - x] \cdot \theta^2} < 0 \tag{2.106}$$

where we used the condition $r = f'(\hat{k}) - \delta$. The expressions in parentheses in equation (2.106) are each positive because $f'(\hat{k}) - \delta$ exceeds $\rho + \theta x$, the steady-state interest rate, which exceeds $n + x$ from the transversality condition. Therefore, $\dot{\gamma}_{\hat{k}} < 0$ follows, Q.E.D.

2.12 Problems

2.1 Preclusion of borrowing in the Ramsey model. Consider the household optimization problem in the Ramsey model. How do the results change if consumers are not allowed to borrow, only to save?

2.2 Irreversibility of investment in the Ramsey model. Suppose that the economy begins with $\hat{k}(0) > \hat{k}^*$. How does the transition path differ depending on whether capital is reversible (convertible back into consumables on a one-to-one basis) or irreversible?

2.3 Exponential utility. Assume that infinite-horizon households maximize a utility function of the form of equation (2.1), where $u(c)$ is now given by the exponential form,

$$u(c) = -(1/\theta) \cdot e^{-\theta c}$$

where $\theta > 0$. The behavior of firms is the same as in the Ramsey model, with zero technological progress.

a. Relate θ to the concavity of the utility function and to the desire to smooth consumption over time. Compute the intertemporal elasticity of substitution. How does it relate to the level of per capita consumption, c?

b. Find the first-order conditions for a representative household with preferences given by this form of $u(c)$.

c. Combine the first-order conditions for the representative household with those of firms to describe the behavior of $\hat{c}$ and $\hat{k}$ over time. [Assume that $\hat{k}(0)$ is below its steady-state value.]

d. How does the transition depend on the parameter θ? Compare this result with the one in the model discussed in the text.

2.4 Stone–Geary preferences. Assume that the usual conditions of the Ramsey model hold, except that the representative household's instantaneous utility function is modified from equation (2.10) to the Stone–Geary form:

$$u(c) = \frac{(c - \bar{c})^{1-\theta} - 1}{1 - \theta}$$

where $\bar{c} \geq 0$ represents the subsistence level of per capita consumption.

a. What is the intertemporal elasticity of substitution for the new form of the utility function? If $\bar{c} > 0$, how does the elasticity change as c rises?

b. How does the revised formulation for utility alter the expression for consumption growth in equation (2.9)? Provide some intuition on the new result.

c. How does the modification of utility affect the steady-state values $\hat{k}^*$ and $\hat{c}^*$?

d. What kinds of changes are likely to arise for the transitional dynamics of $\hat{k}$ and $\hat{c}$ and, hence, for the rate of convergence? (This revised system requires numerical methods to generate exact results.)

2.5 End-of-the-world model. Suppose that everyone knows that the world will end deterministically at time $T > 0$. We worked out this problem in the text when we discussed the importance of the transversality condition. Go through the analysis here in the following steps:

a. How does this modification affect the transition equations for $\hat{k}$ and $\hat{c}$ in equations (2.24) and (2.25)?

b. How does the modification affect the transversality condition?

c. Use figure 2.1 to describe the new transition path for the economy.

d. As T gets larger, how does the new transition path relate to the one shown in figure 2.1? What happens as T approaches infinity?

2.6 Land in the Ramsey model. Suppose that production involves labor, L, capital, K, and land, Λ, in the form of a constant-returns, CES function:

$$Y = A \cdot [a \cdot (K^{\alpha} L^{1-\alpha})^{\psi} + (1 - a) \cdot \Lambda^{\psi}]^{1/\psi}$$

where $A > 0$, $a > 0$, $0 < \alpha < 1$, and $\psi < 1$. Technological progress is absent, and L grows at the constant rate $n > 0$. The quantity of land, Λ, is fixed. Depreciation is 0. Income now includes rent on land, as well as the payments to capital and labor.

a. Show that the competitive payments to factors again exhaust the total output.

b. Under what conditions on ψ is the level of per capita output, y, constant in the steady state? Under what conditions does y decline steadily in the long run? What do the results suggest about the role of a fixed factor like land in the growth process?

2.7 Alternative institutional environments. We worked out the Ramsey model in detail for an environment of competitive households and firms.

a. Show that the results are the same if households carry out the production directly and use family members as workers.

b. Assume that a social planner's preferences are the same as those of the representative household in the model that we worked out. Show that if the planner can dictate the choices of consumption over time, the results are the same as those in the model with competitive households and firms. What does this result imply about the Pareto optimality of the decentralized outcomes?

2.8 Money and inflation in the Ramsey model (based on Sidrauski, 1967; Brock, 1975; and Fischer, 1979). Assume that the government issues fiat money. The stock of money, M, is denoted in dollars and grows at the rate μ, which may vary over time. New money arrives as lump-sum transfers to households. Households may now hold assets in the form of claims on capital, money, and internal loans. Household utility is still given by equation (2.1), except that $u(c)$ is replaced by $u(c, m)$, where $m \equiv M/PL$ is real cash balances per person and P is the price level (dollars per unit of goods). The partial derivatives of the utility function are $u_c > 0$ and $u_m > 0$. The inflation rate is denoted by $\pi \equiv \dot{P}/P$. Population grows at the rate n. The production side of the economy is the same as in the standard Ramsey model, with no technological progress.

a. What is the representative household's budget constraint?

b. What are the first-order conditions associated with the choices of c and m?

c. Suppose that μ is constant in the long run and that m is constant in the steady state. How does a change in the long-run value of μ affect the steady-state values of c, k, and y? How does this change affect the steady-state values of π and m? How does it affect the attained utility, $u(c, m)$, in the steady state? What long-run value of μ would be optimally chosen in this model?

d. Assume now that $u(c, m)$ is a separable function of c and m. In this case, how does the path of μ affect the transition path of c, k, and y?

2.9 Fiscal policy in the Ramsey model (based on Barro, 1974, and McCallum, 1984). Consider the standard Ramsey model with infinite-horizon households, preferences given by equations (2.1) and (2.10), population growth at rate n, a neoclassical production function, and technological progress at rate x. The government now purchases goods and services in the quantity G, imposes lump-sum taxes in the amount T, and has outstanding the quantity B of government bonds. The quantities G, T, and B—which can vary over time—are all measured in units of goods, and B starts at a given value, $B(0)$. Bonds are of infinitesimal maturity, pay the interest rate r, and are viewed by individual households as perfect substitutes for claims on capital or internal loans. (Assume that the government never defaults on its debts.) The government may provide public services that relate to the path of G, but the path of G is held fixed in this problem.

a. What is the government's budget constraint?

b. What is the representative household's budget constraint?

c. Does the household still adhere to the first-order optimization condition for the growth rate of c, as described in equation (2.9)?

d. What is the transversality condition and how does it relate to the behavior of B in the long run? What does this condition mean?

e. How do differences in $B(0)$ or in the path of B and T affect the transitional dynamics and steady-state values of the variables c, k, y, and r? (If there are no effects, the model exhibits *Ricardian equivalence.*)