Springer Texts in Business and Economics

Sibabrata Das
Alex Mourmouras
Peter Rangazas

Economic Growth and Development

A Dynamic Dual Economy Approach

Second Edition

Springer

Springer Texts in Business and Economics

Springer Texts in Business and Economics (STBE) delivers high-quality instructional content for undergraduates and graduates in all areas of Business/Management Science and Economics. The series is comprised of self-contained books with a broad and comprehensive coverage that are suitable for class as well as for individual self-study. All texts are authored by established experts in their fields and offer a solid methodological background, often accompanied by problems and exercises.

More information about this series at http://www.springer.com/series/10099

Sibabrata Das • Alex Mourmouras
Peter Rangazas

Economic Growth and Development

A Dynamic Dual Economy Approach

Second Edition

Springer

Sibabrata Das
Strategy, Policy & Review Department
International Monetary Fund
Washington, DC, USA

Alex Mourmouras
International Monetary Fund
Washington, DC, USA

Peter Rangazas
Department of Economics
Indiana University-Purdue University
Indianapolis, IN, USA

ISSN 2192-4333 ISSN 2192-4341 (electronic)
Springer Texts in Business and Economics
ISBN 978-3-319-89754-7 ISBN 978-3-319-89755-4 (eBook)
https://doi.org/10.1007/978-3-319-89755-4

Library of Congress Control Number: 2018944452

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The second edition improves on the first edition in the usual ways: correcting errors, making the notation more consistent, updating studies, and rewriting the rough passages. Larger changes were also made to allow students with only an elementary background in economics and mathematics to join in. There are two entirely new chapters that unpack the material and slow down the exposition of the extended one-sector growth model. Chapter 3 details the role of fiscal policy in development. Chapter 4 focuses exclusively on schooling and fertility. Over 90 additional exercises are included throughout the book to help build understanding. There is now a technical appendix with examples of how math is used in constructing the economic models. Both the text and the solution manual include more diagrams to illustrate important points.

We had lots of help in making a better book. Undergraduate and graduate students in our development and macroeconomics classes provided many helpful comments. Stephen Rangazas prepared all the new diagrams. Our editor, Lorraine Klimowich, made a convincing case that the second edition was worthwhile. Thanks to all.

Washington, DC, USA — Sibabrata Das
Washington, DC, USA — Alex Mourmouras
Indianapolis, IN, USA — Peter Rangazas

Contents

1 Overview

Among the most enduring questions in economics are those related to growth and development. Since 1776, when Adam Smith published *An Inquiry into the Nature and Causes of the Wealth of Nations*, economists have been assessing the factors determining a nation's standard of living. At Adam Smith's time the focus was primarily on the *level* of living standards. There was much less motivation to study ongoing economic progress because not much sustained growth had ever occurred! The best growth theorist among the classical economists was Thomas Malthus, who explained why living standards *failed* to improve over time and instead remained stagnant in the long-run. The Industrial Revolution was just underway in England and had yet to spread widely across the globe. The standard of living for the vast majority of people was very low. Income disparities across regions of the world were relatively minor. The Western Offshoots (Australia, Canada, New Zealand, and the United States) were the richest countries and African countries were the poorest, but the gap in per capita income was only 3 to 1.

After 1800 the Industrial Revolution and sustained modern growth began to take hold in selected places of the world causing their living standard to diverge from the rest. By the end of the twentieth century, per capita income was 18 times higher in the Western Offshoots than in Africa (Galor (2011, Chapter 1)). Today, countries among the richest 5% have per capita incomes that are at least 25 times that of countries in the poorest 5% (Jones and Vollrath (2013, Table 1.1)). Explaining the huge income gaps across the world has now become *the* question in economics. Nobel Prize winning economist Robert Lucas (2002, p. 21) shifted his attention from business cycle research to economic development because, as he put it,

> 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.

More than two centuries of research studying modern growth have improved our understanding of how investments in physical and human capital, advancing technologies, openness, and sound institutions transform relatively poor economies

S. Das et al., *Economic Growth and Development*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-319-89755-4_1

into economic powerhouses. This book is an introduction to some of the newer features of growth theory that developed after 1950, when research on the topic exploded. We show how the theory can be blended with historical data and case studies to think about the sources of economic prosperity that may be used to help lagging countries also experience prosperity.

Our coverage is selective and the book is by no means a broad survey. We concentrate on transitional growth from a two sector perspective. Most economists believe that this is the right approach for studying early development. However, we also believe that the importance of transitional growth in explaining the complete growth experience of countries over very long periods of time has been underestimated. One of the primary objectives of the book is to make a case for transitional growth and its implications. For these reasons we do not cover *endogenous growth theory*, which is an attempt to explain long-run technological change in already developed economies. For those who have an interest in endogenous growth theory, Aghion and Howitt (2008) is an excellent text for advanced undergraduates and beginning graduate students that discusses the topic in detail.

We begin with a single, relatively simple, theoretical framework that augments the Solow model to include endogenous theories of saving, fertility, human capital, and policy formation. The analysis is then extended to include two sectors of production. We study the structural transformation of developing economies as they shift from traditional production in largely rural areas to modern production in largely urban areas during the early take-off period of modern growth. The two-sector growth model is used to explain the commonly observed differences in saving, worker productivity, and fertility across rural and urban sectors. We examine the effects of policies that reallocate resources across these sectors, such as taxation, migration restrictions, international trade, and an urban bias in the provision of public services. How policies affect the pace of the structural transformation is a critical feature of development as it plays an important causal role in determining an economy's aggregate growth rate.

The extensions to the standard one sector growth model mentioned above add significant complexity to the analysis. We maintain tractability by using specific functional forms that make the main points more transparent. The use of specific functional forms also allows us to calibrate the models and assess the quantitative importance of various sources and mechanisms of growth. We believe this approach makes the book suitable for advanced undergraduates, beginning graduate students, and policy makers specializing in the macroeconomic analysis of development. The mathematics used in most of the text requires only undergraduate calculus and an exposure to optimization theory that can be found in intermediate microeconomics or an undergraduate mathematics-for-economist course. The more technical material is contained in the second half of the book and in the appendices at the end of chapters. An appendix at the end of the book provides mathematical background for those who want to follow the derivations of key equations. A large number of end-of-chapter *Exercises* are offered; *Questions* that help the reader focus on the main points and *Problems* for students who want to work on model building skills or conduct numerical examples that illustrate how the models operate in a concrete way. Solutions for all the *Exercises* are available at the Springer website for the book.

1.1 Beyond the Solow Model

Early post World War II thinking about growth and development concentrated on industrialization or the accumulation of private capital. Private capital accumulation can be analyzed using the famous Solow growth model, the dominant analytical device appearing in undergraduate texts on macroeconomics and development to this day. However, over the latter portion of the twentieth century, academics and policy makers realized that private capital accumulation can only explain a relatively small portion of the growth in worker productivity and living standards. Other sources of growth began to receive attention, such as human capital, fertility, resource allocation, and technological progress. Textbooks have been slow to respond to this realization as the Solow growth model remains the only fully articulated theory of growth provided to students. Apart from the lack of attention paid to other sources of growth, the Solow model is an unsatisfactory tool for studying private physical capital accumulation because it lacks the complete microeconomic foundations needed to link policies to investment and assess welfare effects.

We go beyond the Solow model in two ways. First, we build in the needed microeconomic foundations by using the *overlapping generations* framework developed by Peter Diamond (1965) that has become one of the workhorses of macroeconomic analysis. The overlapping generation model incorporates explicit households that make life-cycle saving decisions. The life-cycle theory of saving allows links to be made between policies and investment in private physical capital. In particular, we include endogenous theories of fiscal policy, including taxation and the accumulation of *public* capital or infrastructure such as roads, public education, and property right protections. Second, we gradually extend the overlapping-generations framework to include household decisions about human capital investments, fertility, and the locational choice for work that affects the efficiency of resource allocation. These extensions allow us to provide an explicit theory with a balanced emphasis on a variety of growth determinants.

1.2 Foreign Aid

A *Great Divergence* in living standards began to form in the nineteenth century. Two centuries later the gaps in per capita income between rich and poor countries are huge. The dramatic gaps are concerning and raise a series of questions. Why are some countries so rich and others so poor? How did the rich countries reach their current level of economic development? What can be done to promote growth in poor countries and help narrow worldwide income inequality?

Most texts in economic growth and development focus on the first two questions but leave the third question to be addressed through recommended readings for students who want to pursue the topic after the course is completed. We think that foreign development aid is too important, and of too much interest to students, to be treated as an afterthought.

The cross-country income gaps of the mid-twentieth century caused international economic assistance to become a prominent feature of the global system since the 1950s. National governments in economically advanced countries created international organizations to provide loans and grants to developing nations in need due to mismanagement, conflict, natural disasters and other bad luck. The most prominent of these international financial institutions (IFIs) are the International Monetary Fund (IMF) and the International Bank for Reconstruction and Development (IBRD), which is commonly known as the World Bank. These two US-based IFIs were created in the 1940s. The mission of the World Bank in particular is to provide assistance to poor countries with the goal of jump-starting sustained economic growth.

Unfortunately, foreign aid has a disappointing track record. The correlation between aid inflows to a poor country and its subsequent economic growth is, at best, weak. International aid organizations are under constant criticism from conservatives and liberals alike. With so much human suffering caused by persistent poverty, frustration with the failed attempts to help developing countries is high. Even more modest attempts to temporarily relieve hunger and illness with humanitarian aid often end in failure.

We devote almost an entire chapter to foreign aid. We attempt to explain why aid has generally not worked and provide suggestions for how international assistance might be improved. More generally, we examine what policy recommendations have the best chance of increasing growth and also discuss the political economy of why these policies are nevertheless resisted.

1.3 Why a Two-Sector Approach?

About half of our text is devoted to the analysis of a two-sector growth model. Given the predominance of the one-sector models in growth theory it is natural to ask if a two-sector model is worth the trouble. As mentioned above, many economists believe accounting for two-sectors is essential, especially for understanding the early stages of development. Lewis (1954) noted that developing economies exhibit a *dualism*, where two economies with fundamentally different structures operate within a single country. One economy operates in a *traditional* sector using elementary production technologies that rely heavily on raw labor, natural resources, and land. The other economy operates in a *modern* sector using advanced technologies that rely heavily on skilled labor and physical capital. The precise interpretation of the two sectors is left open and depends on the particular application of the framework.

In some applications, the traditional sector is thought of as rural in location and the goods produced are assumed to be agricultural. The modern sector is urban in location and produces manufacturing goods. In other applications, the same goods are produced in each sector but using different production technologies and inputs. For example, agricultural products can be produced using traditional or modern methods. As the economy evolves, the traditional ways of producing disappear, but not the products themselves.

The existence and effectiveness of markets also may differ across the two sectors as the markets for labor and capital in the modern sector will generally be more developed than the markets for labor and land in the traditional sector. Under this interpretation, the decline in the traditional sector represents the spread of markets for labor and land (as land is enhanced and developed, essentially becoming part of the reproducible capital stock).

In most developing countries, these different interpretations of the two sector framework strongly overlap. The traditional sector is predominately rural, agricultural, and is operated without much reliance on formal markets for labor and land. In the early stages of growth, the traditional sector is very large, and this is where the dual economy approach has the greatest potential to improve our understanding of development.

When the two sectors are given an explicit geographic interpretation, the households living and producing in the two locations may differ in their behavior because of differences in their economic environments and their initial conditions. Household behavior that may differ across sectors includes saving, work effort, human capital investment, and fertility—all of which relate to aggregate economic growth in important ways.

For example, Lewis thought that a dual approach was necessary to explain why saving rates and capital accumulation increase over the course of development. He conjectured that the income of capital owners in the modern sector would rise relative to incomes of workers and land owners as "surplus" labor from the traditional sector is pulled into the modern sector with little upward pressure on wages. Lewis believed that the relative expansion of capital income was important for growth because capital owners were viewed as saving a larger fraction of their income than land owners and workers. Thus, growth was accelerated by an increase in the economy's saving rate as the modern sector expanded and the traditional sector contracted.

Carter et al. (2003) argue saving rates expand with development for a different reason, but one that is also related to the presence of a dual economy. Residual income from inherited farms finances the consumption of the elderly and reduces the need for retirement saving during working years. As the economy goes through the structural transformation away from traditional family farming, the reliance on income from inherited family farms declines and the retirement saving out of earnings rises.

The difference in saving rates across the two sectors is only one possible feature of dual economies that affects aggregate growth. There is now a substantial body of evidence suggesting that there are large gaps in worker productivity across sectors in the early stages of development (Gollin, Parente, and Rogerson (2002, 2004) and Gollin, Lagakos, and Waugh (2014)). These productivity gaps suggest that labor may be inefficiently allocated and that TFP and aggregate economic growth increase as labor migrates from the low productivity traditional sector to the high productivity modern sector.

It is commonly observed that fertility is much higher in the traditional sector than in the modern sector (e.g. Greenwood and Seshadri (2002)). This fact suggests that the movement of households from the traditional sector to the modern sector lowers the economy's fertility rate. A reduction in the economy's fertility creates another link between the transformation of the dual economy and economic growth. Reductions in population growth allow for greater accumulation in physical capital *per worker*, for a given saving rate.

Years of schooling for children of households in the modern sector are typically higher than those in the traditional sector (Cordoba and Ripoll (2006) and Vollrath (2009)). There is some suggestion that the "quality" of schooling is also different across the two sectors, as students in the rural schools of developing countries are less well equipped and have fewer days of attendance over the course of a school year (e.g. Banerjee and Duflo (2011, Chapter 4)). Thus, growth of the modern sector may increase human capital, yet another important cause of economic growth. Recent theories connect the rise in schooling and the decline in fertility associated with development (Galor (2005, 2011)). Some of the theories relate both behaviors to the dual structure of economies (Greenwood and Seshadri (2002), Doepke (2004), Cordoba and Ripoll (2006), and Mourmouras and Rangazas (2009a)).

A country's fiscal policy is also affected by its dual structure. Countries with larger traditional sectors have a more difficult time collecting taxes and providing essential public infrastructure to private producers (Mourmouras and Rangazas (2009b)). Similar to private capital accumulation, the structural transformation of an economy away from the traditional sector and toward the modern sector may accelerate the growth in productive public capital per worker. There are also many political economy issues associated with how policies are affected by the relative influence of landowners who dominate the traditional sector versus capitalists who dominate the modern sector (e.g. Galor et al. (2009)) or rural versus urban households, who may receive different levels of attention from government officials (e.g. Mourmouras and Rangazas (2013)).

A final connection between the dual economy and aggregate economic growth focuses on the role of cities in the urban sector (Henderson 2010). There are theories and evidence supporting the idea that producing in larger cities can raise worker productivity through knowledge and information spillovers. These externalities are believed to be positively associated with the population density of the city, at least up to the point where various negative effects of crowding begin to dominate. The premise is that the more people that work in a concentrated area, the greater the flow of ideas and the better the match between employers and employees. Raising the stock of knowledge, and matching skills and tasks more effectively, raises worker productivity in the city. Thus, the concentration of workers in larger cities may increase economic growth apart from the other mechanisms described above.

1.4 The Dual Economy

The recent literature on two-sector models and the structural transformation relates back to the earlier work on dual economies, although there are substantial differences between the older and more modern approaches. Lewis (1954) and Ranis and Fei (1961) pioneered the analytical treatment of dual economies and the structural transformation. Important assumptions in their analysis included: (i) an exogenous institutional wage in the agricultural sector and (ii) the institutional wage is paid out of the average product of labor (which includes land rents) and (iii) a "surplus" of labor in the agricultural sector with the marginal product of labor in agriculture lower than the marginal product of labor in industry (which equals the institutional wage and average product of labor in agriculture). There are no wage gaps in their models, as the institutional wage also determines the wage paid to labor in industry. In the early stages of development, with most of the labor in traditional agriculture, the marginal product of labor in agriculture was thought to be, not just low, but actually zero. This was Lewis's extreme interpretation of "surplus" labor.

The empirical relevance of surplus labor, even in the poorest developing countries, was challenged early on by Schultz (1964). He used the natural experiment of an influenza epidemic in India as a test of the idea. The epidemic caused a sharp decline in the agricultural labor supply and total agricultural output fell accordingly—a clear contradiction to the surplus labor assumption. The surplus labor assumption certainly is not relevant to labor scarce countries, such as the U.S. in the nineteenth century, because migrant workers from other sectors of the economy were typically hired during harvest times.

The assumption of an exogenous institutional wage that determines wages across the economy is also problematic. As mentioned above, not only are there large productivity gaps across sectors, but there are large annual wage gaps as well. Explaining the presence of these large wage gaps has become the main focus of the recent literature on dual economies.

The earlier dual theories were naturally not unified in the sense that models are today. They ignored human capital altogether. The theories did recognize the importance of population growth and physical capital formation for the economic transformation, but these variables were treated exogenously. Schultz (1964) was the first to stress the importance of human capital in causing the movement away from traditional agriculture. Eaton (1987) and Drazen and Eckstein (1988) modeled the interaction between a dual structure and physical capital formation. Galor and Weil (1996, 2000) insisted that fertility and population growth must be given a central role in theories of long-run development. This book incorporates all these features from the recent literature into a dual economy approach.

To see how a unified approach to modeling the dual economy is able to generate connections between key features of development, consider the treatment of implicit claims on land. In the older theories, there were implicit claims on land rents that allowed "wages" to exceed the marginal product of labor in the traditional sector, resulting in an inefficient allocation of labor. The older theories were static model, with no markets for land, and the total output was simply assumed to be split across the workers.

In the newer theories, inefficiencies in the allocation of labor are also linked to land rents. The current generation of landowners desires to bequeath land to the next generation if they are willing and able to maintain the tradition of family farming (a preference that is more likely to be operative when land markets are incomplete). Workers in the traditional sector accept lower wages (which do equal their marginal product) in the traditional sector because if they move off the farm they lose their claims to land in the future, i.e. they will not inherit the land from their parents or tribal elders. Thus, expectation of future rents from an intergenerational transfer of land creates both wage and productivity gaps across sectors. Moreover, the expectation of future land rents also lowers saving and raises fertility. The observed sector differences in wages, productivity, saving, and fertility are all explained in the same way.

1.5 Growth Facts

The focus and organization of the book is guided by the goal of explaining key stylized growth facts. Explanations of these facts are developed throughout the book. The facts are listed roughly in the order in which they are addressed.

G1—The marginal product of capital and the rates of returns to assets are modestly higher at low levels of development.

G2—Children spend more time in school, within and across years, and less time working as an economy develops.

G3—Population growth rates rise and then fall as economies develop.

G4—The per capita income gaps between rich and poor countries today are huge.

G5—Standard measures of physical and human capital accumulation are not sufficient to fully account for the growth within countries or the income differences across countries.

G6—After 1950, some poorer countries experienced Growth Miracles that were not witnessed before in human history, while other poor countries experienced Growth Disasters that lowered their living standards.

G7—There was almost no sustained growth in living standards before 1800.

G8—The onset of modern growth began in some countries around 1800, but was delayed in others, creating the Great Divergence.

G9—As an economy develops, there is a structural transformation away from home-based informal production to firm-based formal production. A shift of labor out of agriculture and a movement of the population from rural to urban areas are typically associated with the structural transformation.

G10—The shares of household budgets devoted to food fall over the course of development and caloric intake stays relatively constant for centuries.

G11—The level of real wages and labor productivity is much lower in the agricultural sector than in other sectors of the economy during the early stages of development.

G12—The relative size of government grows as economies develop (Wagner's Law). The government shares of currently poor countries are larger than the historical shares of currently rich countries when at similar stages of development.

1.6 Outline

To summarize, the key features of the book are (i) an extension of the standard model of private physical capital accumulation that provides a more balanced theory of the determinants of growth, (ii) a focus on foreign aid and an examination of pro-growth policies, and (iii) a two-sector approach that connects the structural transformation, the demographic transition, and economic growth. The book is outlined as follows.

Chapter 2 presents a one-sector model of physical capital accumulation where saving is motivated by the need to finance retirement consumption, based on the seminal work of Diamond (1965). The model is then extended to include intergenerational transfers that take the form of both physical assets and investment in children's human capital. Calibration exercises show that one can explain very little growth in U.S. history unless human capital is included. This result receives further support when we move to a two sector setting in the second half of the book.

Chapter 3 adds the government. In particular, we include the aspects of fiscal policy that affect growth such as taxation and public infrastructure. Taxation reduces private saving and capital accumulation. However, taxes also fund public infrastructure investments by the government—a vital ingredient to sustained growth. We focus on how to balance the interaction between these two features of fiscal policy.

Chapter 4 moves beyond physical capital accumulation to examine the roles of fertility and human capital. Based on the famous theory of Gary Becker, households are modeled as choosing the "quantity" and "quality" of children. At the economy-wide level, the household choices determine population growth and human capital formation. The chapter also presents Malthus's theory of how population growth prevents long-run economic progress.

Chapter 5 includes all the elements from Chaps. 2, 3, and 4, providing a complete one-sector growth model. Large income differences across countries are explained by a combination of a *poverty trap*, that keeps human capital low and fertility high, and *anti-growth fiscal policies*. The model is extended to include foreign aid and pro-growth policy recommendations, including a discussion of their limitations.

Chapter 6 provides an introduction to a two sector model with complete markets across the economy. We use the model to examine the origins of modern growth, asset bubbles, how the structural transformation affects physical capital accumulation , the impact on economic growth and welfare of opening the economy to the trade of goods, and how health considerations help explain the relatively constancy of caloric intake and the decline in household budget shares spent on food over the course of development.

Chapters 7, 8, and 9 introduce market imperfections and cultural attitudes that create the dualism observed in many developing countries. The dual structure leads to differences in household behavior and to policy conflicts. Chapter 7 focuses on gaps in worker productivity and fertility. Chapter 8 shows how differences in saving behavior and the internal migration of workers affect physical capital accumulation per worker in the modern sector. Chapter 9 combines the material in Chaps. 5 and 6 to provide a complete dynamic analysis of a dual economy. The model is used to simulate transitional growth over two centuries. The simulated growth and predictions about other features of development are compared to real world data.

Chapter 10 introduces fiscal policy in a two-sector setting to examine the connection between government and urbanization. We find that it can be efficient for fiscal policy to be "biased" toward the modern sector, as the bias encourages internal migration that raises worker productivity throughout the economy. However, the rural sector cannot be completely ignored. Some rural development is needed to maintain a reasonable pace of migration to the city in order to prevent crowding of government services. We also find that internal migration restrictions is not an efficient way to pace the structural transformation as they make modern urban sector households better off at the expense of rural households.

Chapter 11 provides a conclusion.

The book can provide an undergraduate course on economic growth and development by moving at a deliberate pace through Chaps. 2, 3, 4, and 5 for a thorough coverage of the extended one-sector model. If one wants to avoid the complexity and policy emphasis of Chap. 5, the beginning sections of Chaps. 6 and 7 can be used instead to offer an exposure to the two-sector approach. Masters and beginning PhD courses on economic growth can use Chaps. 6, 7, 8, 9, and 10 to provide a dual economy perspective.

References

Aghion P, Howitt P (2008) The economics of growth. Boston: MIT Press.

Banerjee A, Duflo E (2011) Poor economics. Public Affairs, New York.

Carter S, Ransom R, Sutch R (2003) Family matters: The life-cycle transition and the unparalleled fertility decline in antebellum America. In: Guinnane, T Sundstrom W, Whately W (eds) History matters: Essays on economic growth, technology, and demographic change. Stanford University Press, Stanford, California.

Cordoba J, Ripoll M (2006) Agriculture, aggregation, wage caps, and cross-country income differences. Working paper 246, University of Pittsburgh Department of Economics.

Diamond P (1965) National debt in a neoclassical growth model. American economic review 55 (5):1126–1150.

Doepke M (2004) Accounting for fertility decline during the transition to growth. Journal of economic growth 9(3):347–383.

Drazen A, Eckstein Z (1988) On the organization of rural markets and the process of development. American economic review 78(3):431–43.

Eaton J (1987) A dynamic specific factors model of international trade. Review of Economic Studies 54:325–338.
Galor O (2005) From stagnation to growth: Unified growth theory. In: Aghion P, Durlauf S (eds) Handbook of economic growth. Amsterdam, North Holland.
Galor O (2011) Unified growth theory. Princeton University Press, Princeton, New Jersey.
Galor O, Moav O, Vollrath D (2009) Inequality in land ownership, the emergence of human capital promoting institutions, and the great divergence. Review of economic studies 76(1):143–179.
Galor O, Weil D (1996) The gender gap, fertility, and growth. American economic review 86: 374–387.
Galor O, Weil D (2000) Population, technology, and growth: From Malthusian stagnation to modern economic growth. American economic review 90(4):806–828.
Gollin D, Lagakos D, Waugh M (2014) The agricultural productivity gap. Quarterly journal of economics 129(2):939–993.
Gollin D, Parente S, Rogerson R (2002) The role of agriculture in development. American economic review 92(2:160–164.
Gollin D, Parente S, Rogerson R (2004) Farmwork, homework, and international income differences. Review of economic dynamics 7(4):827–50.
Greenwood J, Seshadri A (2002) The US demographic transition. American economic review. 92 (2):153–59.
Henderson V (2010) Cities and development. J Reg Sci 50(1):515–540.
Jones, C and Vollrath, D (2013) Introduction to economic growth, New York: W.W. Norton.
Lewis A (1954) Development with unlimited supplies of labor. The Manchester School 22:139–92.
Lucas, R (2002) Lectures on economic growth. Cambridge, MA: Harvard University Press.
Mourmouras A, Rangazas P (2009a) Reconciling Kuznets and Habbakuk in a unified growth model. Journal of economic growth 14(2):149–181.
Mourmouras A, Rangazas P (2009b) Fiscal policy and economic development. Macroeconomic dynamics 13(4):450–476.
Mourmouras A, Rangazas P (2013) Efficient urban bias. Journal of economic geography 13 (3):451–471.
Ranis G, Fei J (1961) A theory of economic development. American economic review 51 (4):533–565.
Schultz T (1964) Transforming traditional agriculture. New Haven: Yale University Press.
Vollrath D (2009) How important are dual economy effects for aggregate productivity? Journal of development economics 88(2):325–334.

Part I

One-Sector Growth Models

2 Overlapping-Generations Model of Economic Growth

This chapter introduces the one-sector neoclassical growth model with overlapping generations. The primary focus of the chapter is growth via private physical capital accumulation. We think of *private physical capital* as manmade durable inputs to the production process. For our purposes, private capital can be primarily thought of as plant and equipment that is produced in one period and then used in production in the following period.[1] To model production, we introduce *firms*, economic institutions that combine physical capital and labor to produce goods and services. Later in the chapter, we introduce *human capital*, the knowledge and skills embodied in workers. Chapter 3 adds *public capital*, the economy's infrastructure created by the government, such as roads, laws, and utilities.

The accumulation of capital must be financed or funded by household saving. We use the two-period *life-cycle theory* of household consumption as the basis for explaining saving behavior. In the overlapping-generations model, households save during their working period to finance retirement consumption. This implies that, instead of a *single representative* household type, there are *two different household types representing two different generations that overlap* each period—a "young" working household and an "old" retired household.

Once the theoretical model is developed, we apply it to real world issues. First, we illustrate how the model can be "estimated," or more precisely *calibrated*, to make quantitative analysis possible. We then examine how well the simple model of physical capital accumulation can replicate the economic growth in the U.S. from 1870 to 2000. The main finding is that physical capital alone can only explain a small portion of the observed growth. Human capital is added and significantly improves the model's ability to match the historical facts.

[1]Definitions of physical capital will vary depending on the purpose at hand. In some cases physical capital is defined to include inventories, software, land, and other inputs that extend beyond plant and equipment. Public capital is often lumped together with private capital as if they are close substitutes. Chapter 3 argues that it makes more sense to treat them as complementary inputs.

S. Das et al., *Economic Growth and Development*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-319-89755-4_2

The chapter will explain *G1*, the fact that returns to capital are modestly higher at the early stages of development. It will also become clear that even the combination of physical and human capital accumulation is not sufficient to explain all of an economy's growth, *G5*.

2.1 Firms, Production, and the Demand for Capital

The first step in developing a general equilibrium model of output and income is to introduce a production technology. We assume that production takes place in "firms"— organizations that hire labor and rent capital in order to produce output. Each firm's production knowledge or "technology" is represented by a *Cobb-Douglas production function*,

$$Y_t = AK_t^{\alpha} M_t^{1-\alpha}, \tag{2.1}$$

where Y denotes output, K denotes the capital stock rented, M denotes the hours of work hired, and where A and α are technological parameters. The production function is a technological "recipe" that relates the inputs hired and used by the firm to the output that the firm is capable of producing. The parameter A is sometimes referred to as *Total Factor Productivity* (TFP). It captures a wide variety of unmeasured variables that affect the productivity of labor and capital; from climate and geography that determine natural resources available and the health environment of households to laws and regulations that restrict the way that production is carried out. The parameter α is a fraction that measures the relative importance of physical capital in the production process. This interpretation of α will become more clear as the theory of the firm is developed below.

The output produced by firms is a single "all-purpose" good that can either be consumed or invested as a physical asset (somewhat like corn that can be either consumed or stored and invested as a physical asset to plant and produce more corn in the future). This abstraction avoids the complication of having two distinct sectors of production, one producing consumer goods and the other capital goods. For some purposes one may require this more elaborate two-sector model, but this is not the way to begin an analysis of a growing economy.

The Cobb-Douglas production function is a special case of what is called a "neoclassical" production function. All neoclassical production functions have three general properties: (i) positive marginal productivity, (ii) diminishing marginal productivity and (iii) constant returns to scale. Economists believe that these properties are common to most production processes.

The marginal product of an input is the increase in output that results from an increase in the use of an input. Formally, it is the *partial* derivative of the production function with respect to a particular input, holding other inputs constant (see the Technical Appendix for a discussion of partial derivatives). For a Cobb-Douglas production function, the *marginal product of labor* and the *marginal product of* *capital* are $(1-\alpha)AK_t^{\alpha}M_t^{-\alpha}$ and $\alpha AK_t^{\alpha-1}M_t^{1-\alpha}$. The marginal productivity of increasing the level of either input is always positive—more output results when the firm hires either more labor or more capital.

Diminishing marginal productivity means the additional output, associated with adding an additional unit of an input, decreases as more of that input is used. While output increases as the firm uses more of an input, the size of the increase gets smaller as the amount of the input used in production increases. Diminishing marginal productivity is based on the intuitive notion of "input crowding." The increasing scarcity of the input held fixed, limits the production that results from adding more of the other input. For example, if there is a given amount of capital, as more workers are hired the amount of capital that each worker can use decreases—serving to limit the rise in output. Note that the marginal product of labor expression above is decreasing in M_t, for a fixed value of K_t. The analogous observation applies to the marginal product of capital. Sketch the marginal product of labor against the level of employment to see this graphically. A similar sketch applies to the relationship between the marginal product of capital and the capital used in production.

Constant returns to scale means that if *both* inputs were increased in the same proportion, then the ability to produce output would also increase by that proportion. This property makes sense because if the firm can simply duplicate its current plant, equipment, and work force, it should be able to duplicate or double its output as well.

Finally, note that the properties we just described imply that the marginal product expressions can be simplified by combining M_t and K_t into the *capital-labor ratio*, also known as *capital intensity*, $k_t \equiv K_t/M_t$. The simplified expressions for the marginal products are, $(1-\alpha)Ak_t^{\alpha}$ and $\alpha Ak_t^{\alpha-1}$. The marginal product of labor is increasing in capital intensity. The more capital per worker, the more productive an additional worker is. The marginal product of capital is decreasing in capital intensity. Higher capital intensity means there are fewer workers available to work with any additional capital bought to the workplace.

The fact that the marginal products of capital and labor are both functions of the capital-labor ratio, k, and not the levels of K and M, is a consequence of the constant returns to scale assumption. This property implies that the *scale* of a firm is indeterminate, i.e. the optimal size of a firm cannot be pinned down by the theory. Firms are indifferent about the level of production, but they do want to hire capital and labor in a particular ratio that depends on the relative market prices of the inputs.

From the point of view of microeconomics, the indeterminacy of firm size can be seen as a disadvantage. One is forced to simply assume that firms are of a given size and that there are enough of them competing to justify the *perfect competition* assumption that is discussed below and used throughout the book. From a macroeconomic point of view, the indeterminate size of firms can be seen as a convenient simplification. The key expressions that characterize the production side of the economy apply to both the individual firm and to the collection of firms as a whole. This is why in many macroeconomic models the distinction between the individual firms and production in the economy as a whole is not emphasized.

What makes (2.1) special in the class of neoclassical production functions is that the Cobb-Douglas functional form implies that the *shares* of national output that are paid to capital owners and workers are the *constant* output elasticity values α and $1-\alpha$. Data shows that over the last century, income shares have in fact stayed roughly constant within and across countries. For this reason, many economists view the Cobb-Douglas functional form as a reasonable approximation to an economy's aggregate production technology. To explicitly see that (2.1) has the constant income share property, we next need to think about how capital owners and workers are paid.

We assume that markets are perfectly competitive in our production economy. As discussed in elementary economics, the notion of competitive markets applies not only to the markets for goods but also to the factor markets for labor and capital. The competitive assumption applied to the factor markets means that firms demand inputs to maximize profits taking as given the market prices of the inputs: the wage rate paid to labor (w) and rental rate on physical capital (r). No single firm is large enough to be able to influence market prices when they unilaterally change their production or input levels. The price of the economy's single output good is taken to be one. So we can think of output and revenue as being the same. Therefore, profit can then be written as $Y_t - w_t M_t - r_t K_t$.

Maximizing profits requires that firms hire capital and labor as long as the marginal benefit (marginal product) exceeds the marginal cost (factor price). Formally, the necessary conditions for profit maximization are

$$\alpha A k_t^{\alpha-1} = r_t \tag{2.2a}$$

$$(1-\alpha) A k_t^{\alpha} = w_t. \tag{2.2b}$$

Equation (2.2a, 2.2b) says that, in order to maximize profit, the marginal product of each input must be equated to its market price, just as in the theory of competitive factor markets from intermediate microeconomics.

From the perspective of an individual firm, that takes factor prices as given, it appears that there are two independent Eqs. (2.2a) and (2.2b), to determine one unknown, k. In general, this situation leads to inconsistent solutions for k—i.e. different solutions for k from each equation. This is not the case here because of an important implication of competitive markets: economic profits are driven to zero. Competition between firms for the available resources will force factor prices to satisfy these equations, which in turn implies that economic profits are zero. Thus, (2.2a) and (2.2b) also play a role in determining the market factor prices and not just k.

To think about this last point further, first notice that we can write the production function as $Y_t = A k_t^{\alpha} M_t$. Next, multiply each side of (2.2a) and (2.2b) by K_t and M_t respectively to get

$$\alpha Y_t = r_t K_t \tag{2.3a}$$

$$(1-\alpha)Y_t = w_t M_t \tag{2.3b}$$

Equation (2.3a) shows that the share of output and revenue paid to owners of capital (by each firm and in the economy as a whole) is the constant, α, an interpretation that was suggested above. Moreover, if αY_t goes to capital owners as a gross rent to capital, there is just enough revenue left over, $(1 - \alpha)Y_t$, to pay workers the competitive wage, implying that economic profit is zero.

The connection made in (2.3a, 2.3b) allows us to refer to α and $1 - \alpha$ as the capital and labor shares. The fact that the Cobb-Douglas technology, combined with competitive markets, implies constant factor shares is a strong prediction of the model. Remarkably, this prediction is approximately consistent with empirical evidence that shows little trend in factor shares as a country develops.

The two Eqs. (2.2a) and (2.2b) are then profit-maximizing conditions that determine two variables: the firm's demand for capital relative to labor and, via the zero profit condition, one of the factor prices. To determine the remaining factor price, we need the final requirement of a competitive equilibrium: *market clearing*. The firm's demand for capital per worker must equal the supply of capital per worker coming from households. We will think of the rental rate on capital as the "price" that clears the capital market. Then interpreting (2.2a) and (2.2b) as determining the demand for capital and the competitive wage rate that generates zero profit, we have three conditions to determine the three unknowns: r_t, w_t, and k_t.

The first step in developing the market clearing condition is to be more explicit about what we mean by the *demand* for capital in the production economy. Start by thinking of the capital-labor ratio on the left-hand side of (2.2a) as the capital-labor ratio *demanded* by firms at different rental rates for capital. Call the firm's demand for k, k_t^d. In period t, firms will enter the capital market to rent capital that they can use in production. Solving (2.2a) for k, we can write the demand for capital in period-t as

$$k_t^d = \left[\frac{\alpha A}{r_t}\right]^{1/(1-\alpha)}. \tag{2.4}$$

Equation (2.4) indicates that as the rental rate required by the market rises, the firm's demand for capital declines. This is because, as the cost of capital rises, firms will shift towards using less capital and more labor in production.

The theory thus far gives us the firms' demand for capital intensity. Now we need to develop a theory for the supply of capital in period t. In other words, we need to discuss who owns the capital and how much capital they are willing to supply to the market.

2.2 Household Saving and the Supply of Capital

In our model, households purchase capital as an asset, a type of saving used to finance retirement consumption. The capital generates funds for retirement consumption purchases when the households rent the capital to firms. So, the supply of capital referred to at the end of Sect. 2.1 results from older households attempting to generate income for retirement consumption.

To capture a retirement motive for saving in the simplest way possible, we assume households live for two periods: one when they are young and working and one when they are old and retired. This means that in any one period there are two household-types from distinct generations: a young working household and an old retired household. Macroeconomic models where different generations operate as distinct decision-makers in each period are called *overlapping generations* models.

Including the saving behavior of households is an important extension to the Solow model of capital accumulation from undergraduate macroeconomic courses. In the Solow model saving is treated as an exogenous variable. The economy's saving rate is simply assumed to be a constant fraction of total income with no explanation provided. In contrast, we derive the saving rate from the utility maximizing behavior of households. When the government is introduced in Chap. 3, the economy's saving rate will be influenced by fiscal policy.

2.2.1 The Supply of Labor and Capital

As just mentioned, the supply of capital that is rented to firms is owned by old retired households. They rent the capital to firms to generate income that finances their retirement consumption. Once the firms complete production using the capital, the retired households sell the capital to the young working households that are looking to save assets to finance their future retirement consumption. The sale of capital provides further resources for retirement consumption of the current old households.

Formally, the currently old households who own and supply the capital, purchased the capital as an asset during their working lives in the previous period. In period $t-1$, each young household supplied one unit of labor to firms and earned the wage, w_{t-1}. With each household supplying one unit of labor, the aggregate supply of labor in each period is then just the number of young households. In period $t-1$, the total supply of labor to all firms is the total number of young households from that generation, $M_{t-1}^{s}=N_{t-1}$.

The capital supplied per unit of labor results from the household's saving behavior, s_{t-1}. Young households save in period $t-1$ by purchasing output and treating it like a physical asset that generates income during retirement by supplying or renting it to firms for use in production during period-t. The firms use this physical capital to produce output and generate revenue in period t. The firms then return the capital, that has been depreciated by use in production, back to households and pay them the rental rate r_t. So, for every unit of capital that households purchase and rent to firms, they receive back in period t, $1-\delta+r_t$, as their return to saving, where δ is the fraction of capital that depreciates from use. We somewhat loosely refer to $r_t-\delta$ as both the "return to capital" and the "interest rate" on household saving.

The total supply of new capital to the market in period-t is the total saving of young households in period $t-1$, $s_{t-1}N_{t-1}$. To match the demand concept in (2.4), we need an expression for the capital supplied *per worker* in period t. The supply of capital per worker in period t is $k_t^s \equiv s_{t-1}N_{t-1}/M_t^s = s_{t-1}N_{t-1}/N_t = s_{t-1}/n$, where n is the average number of children born in each young household. We treat n as an exogenous constant. The number of children each household has determines the relative population size of different generations. For example, if $n=1$, then generations are of equal size and $N_t=N_{t-1}$. If $n>1$, $N_t>N_{t-1}$ and there is positive population growth over time. Note that the *rate* of population growth is $(N_t/N_{t-1})-1=n-1$.

In summary, the factors of production supplied by the households in period t, for hire by firms, are

$$M_t^s = N_t$$

and

$$k_t^s = \frac{s_{t-1}}{n}.$$

To complete the model, we need a theory of s_t.

2.2.2 Household Saving

We now develop a theory of household saving. Households do not directly benefit from saving but rather use saving to create their desired lifetime consumption path. The consumption path that households prefer depends on their attitudes about consuming now rather than later in life. Household preferences are represented by a utility function. The utility function captures the household's preference for consuming at different points in their lifetime.

We assume that household preferences are represented by a time separable, log utility function,

$$U(c_{1t}, c_{2t+1}) = \ln c_{1t} + \beta \ln c_{2t+1}.$$

The *single period* utility function, ln$c$, has the familiar characteristic, one you may recall from introductory economics, of *diminishing marginal utility*. In other words, greater consumption increases satisfaction but at a diminishing rate. All increasing concave functions have this property because their slopes get smaller as the argument of the function increases. The natural log function we are using as our single-period utility function is simply a convenient increasing concave function.

The lifetime utility function includes the satisfaction the household expects to receive from a particular plan for both current and future consumption, combining the utility gained in each period of life. The parameter capturing the household's preferences about the timing of consumption is the pure time discount factor (β). Typically, one assumes that $\beta < 1$ because people are generally viewed as being "impatient," i.e. they weigh utility gained from current consumption higher than utility gained from future consumption.

The household's task is to choose a path for consumption that makes U as large as possible. For a generation-t household, consumption in the first and second periods, c_{1t} and c_{2t+1} determine the value of lifetime utility. Households face constraints that restrict the consumption paths they can afford. In each period there is a budget constraint that must be satisfied. In the first period of life, a generation-t household has its wage (w_t) as a source of funds that can be used to purchase output for consumption (c_{1t}) or for saving (s_t). This gives the first period budget constraint, $c_{1t} + s_t = w_t$. In the second period, consumption (c_{2t+1}) is financed by the saving from the first period, $c_{2t+1} = R_t s_t$, where $R_t = 1 + r_{t+1} - \delta$ is the return from owning physical capital or what sometimes is called the "interest factor." Note the return from saving this period is paid next period based on the rental rate that prevails in that period. In other words the return from s_t is based on r_{t+1}. The *two single period* budget constraints can be combined to form a *single lifetime* budget constraint that requires the present value of consumption to equal the first period wage, $c_{1t} + c_{2t+1}/R_t = w_t$.

Households maximize lifetime utility subject to the lifetime budget constraint. The solution to this problem gives us the optimal consumption and saving behavior of a household

$$c_{1t} = \frac{1}{1+\beta} w_t \tag{2.5a}$$

$$c_{2t+1} = \frac{\beta}{1+\beta} R_t w_t \tag{2.5b}$$

$$s_t = \frac{\beta}{1+\beta} w_t \tag{2.6}$$

All behavior is proportional to the household wage, via an income effect. Higher first period wages make the household want to consume more in *both* periods. The only way it can consume more in the second period is to save some of the greater first period wage income.

2.2.3 Supply of Capital per Worker

Using Eq. (2.6), dated for a generation-t−1, and the definition of k_t^s that was introduced previously, we can now write the economy's supply of capital per worker as

$$k_t^s = \frac{\beta}{1+\beta}\frac{w_{t-1}}{n}. \tag{2.7}$$

The economy's supply of capital per worker next period is based on the saving per worker in the previous period and the growth of the economy's work force. An increase in the previous period's wage raises saving because a portion of the higher wage is consumed and a fraction is put aside to allow consumption in the future to rise as well.

The extent to which saving and capital supplied raises the capital-labor ratio in the next period, depends on the growth in the workforce. Greater fertility implies a higher rate of population growth and a faster growing workforce. As the workforce next period rises relative to the current workforce, less saving and capital will be available per worker in the future. Thus, higher rates of population growth lower the capital-labor ratio by forcing the available capital to be spread over a larger workforce.

2.3 Competitive Equilibrium in a Growing Economy

Before moving to the determination of the market clearing condition in the capital market, let's summarize the key actions taken in period t by each agent.

Firms	hire labor, pay each worker w_t
	rent physical capital per worker, k_t^d, pay owners r_t per unit supplied
Young	
Households	supply one unit of labor, receive w_t
	purchase $s_t = nk_{t+1}^s$ units of physical capital
Old	
Households	supply $s_{t-1} = nk_t^s$ units of physical capital, receive r_t per unit supplied

A market clearing equilibrium in the capital market requires that the firms' demand for capital per worker equals the supply of capital per worker by old households, i.e. $k_t^d = k_t^s$ for all values for t. As in other competitive markets, the market price is the mechanism for bringing the two sides of the market together. In the capital market, the market price is the rental rate on capital that is paid by those

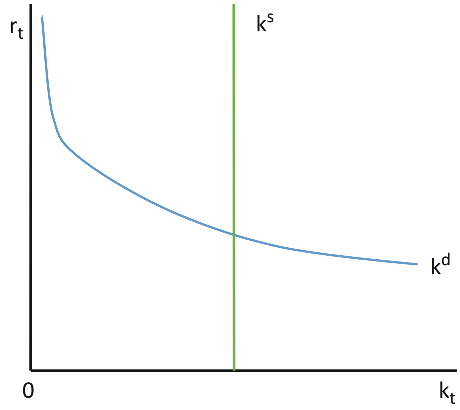


Fig. 2.1 Market clearing equilibrium in the capital market

demanding capital and received by those supplying the capital. Market clearing requires finding a value of r_t that equates (2.4) and (2.7) in every period, as sketched in Fig. 2.1.[2]

Figure 2.1 is the standard demand-equals-supply way of thinking about how equilibrium is determined. It is analogous to the "loanable funds" market of saving and investment commonly used in introductory macroeconomics. Here, the "demand for funds" is replaced by the direct demand for physical capital to be used in production. The "supply of funds" results from household saving. The only way households can save in the model is to directly purchase physical capital and then rent it to firms to use in production. The behavior of firms and households, in demanding and supplying capital in a competitive market, determines the equilibrium return to capital and the amount of capital traded.

While intuitive, the demand-supply approach has limitations as an analytical tool. The problem is that it is a static snapshot of a dynamic economy. In general, a production economy will experience capital accumulation over time. In other words, the k_t determined in the figure will be larger than k_{t-1}. This implies that, using (2.2b), w_t will be larger than w_{t-1}. The increase in wages over time will cause the supply curve in the figure to shift to the right each period. Thus, the diagram reveals that growth in the economy is due to the effect of capital on wages. As the capital stock increases, wages increase. The increase in wages, increases saving and leads to further capital accumulation. However, there are important details of the growth process that are not revealed by this essentially static depiction.

Fortunately there is a nice way of displaying the dynamics of the economy more explicitly. One can substitute the factor price equations from (2.2a, 2.2b), dated for period t−1, into (2.7) and impose the equilibrium condition $k_t^d = k_t^s$ to get

[2]You can think of the value of r_t as actually determined in period $t-1$. In that period households make their saving decision based on the firms' commitments to rent capital in period t and pay the rental rate r_t. In other words, r_t is determined in period t−1 based on the savings behavior of households and the *planned* investment demands of firms.

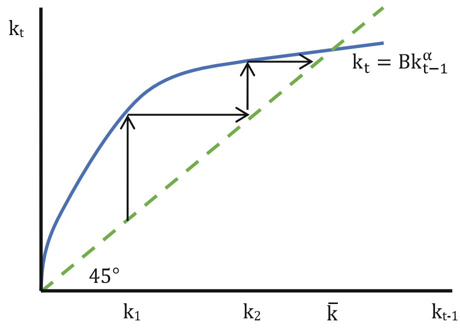


Fig. 2.2 Transitional growth

$$k_t = \frac{\beta}{1+\beta}\frac{(1-\alpha)Ak^{\alpha}_{t-1}}{n} = \mathrm{B}k^{\alpha}_{t-1}, \tag{2.8}$$

$$\text{where } \mathrm{B} \equiv \frac{\beta}{1+\beta}\frac{(1-\alpha)A}{n}.$$

Mathematically, Eq. (2.8) is known as a *difference equation*, which is the discrete-time analog to the differential equation in continuous time that may be more familiar from calculus classes. The difference equation highlights the underlying dynamics of the model that is driven by changes in the capital-labor ratio over time. In economics, Eq. (2.8) is referred to as a *transition equation* because it describes how the economy evolves over time.

The dynamic features of (2.8) can be easily sketched by plotting k_t against k_{t-1} as in Fig. 2.2. Imagine that the economy begins at $k_{t-1} = k_1$. To find out what the capital-labor will be in period 2, move vertically up to the transition equation to find the value of k one period ahead, k_2. In period 2, k_2 will now be the initial capital-labor ratio. To trace the new starting value for k in period 2, move horizontally from the transition equation to the 45-degree line and then back down vertically to the horizontal axis. The process then repeats itself over and over until one reaches $k_t = \bar{k}$, where the transition equation crosses the 45-degree line.[3] At this point, the capital-labor ratio remains constant from period to period and the economy is said to have reached a *steady state* equilibrium. An algebraic solution for the steady state is found by setting $k_t = k_{t-1} = \bar{k}$ in (2.8) and then solving the equation for $\bar{k}$. The transition equation given by (2.8) is simple enough to allow an explicit solution for the steady state capital-labor ratio, $\bar{k} = \mathrm{B}^{\frac{1}{1-\alpha}}$.

The transition diagram reveals an important prediction about economic growth via capital accumulation. In the early stages of growth, period to period changes in k_t are relatively large and the economy grows fast. Over time, the increases in k_t get smaller and the economy's growth rate slows down, until growth ceases altogether in

[3]The economy never literally reaches the steady state, although it will get arbitrarily close.

the steady state. From the static demand and supply figure, we know that growth occurs due to the effect of capital accumulation on wages and saving. What the transition diagram makes clear is that the effect of capital accumulation on wages becomes weaker over time. There is a diminishing effect of k_t on w_t because α is less than one. When an economy is undeveloped and capital is scarce, the creation of new physical capital significantly raises worker productivity and wages. However, as the economy industrializes, the impact of further capital accumulation weakens.[4]

Notice two things about the steady state. First, as k_t grows in approaching the steady state, we know from (2.2a, 2.2b) that interest rates will be falling and wages will be rising. Once the steady state is obtained, because k_t is constant, interest rates and wages must also be constant. Thus, the steady state is characterized by constant interest rates and *zero* growth in labor productivity, real wages, and consumption. In many developed countries, the *average* values of interest rates and returns to capital have been relatively constant over long-periods of time—suggesting that we might view the average position of the economy as being a steady state (with some annual business cycle fluctuations around the economy's typical or average position). However, these same economies are observed to experience *positive* growth rates in labor productivity and real wages *on average*. According to our model, if interest rates show no downward trend, then this positive growth cannot come from increases in the capital-labor ratio. Where does persistent, long-run growth come from after the steady state capital-labor ratio is obtained?

2.3.1 Steady State Growth—Technical Progress

One explanation for persistent economic growth is technical progress—that is increasing knowledge that improves productivity. Technical progress can be thought of as improved production designs or improved factories and equipment. To grow in the steady state with the same amount of capital per worker, we have to get smarter about how we use and design the capital. There are some attempts to explicitly model the research and development process that leads to technical process, but often economists treat technical progress as an exogenous variable, as we do here.

Think of technology as the current stock of *disembodied* blueprints for production methods and machine designs. The state of technology in period t affects the productivity of the workforce. We assume that there is an index number, D_t, that measures the extent to which the state of technology influences the *effective* workforce. The effective workforce in period t is defined as $H_t = D_t M_t = D_t N_t$, which replaces M_t as an input in the Cobb-Douglas production function. When D_t increases, it raises the effective workforce proportionately. For example, if D_t doubles,

[4]The weakening effect of the capital-labor ratio on wages, stems from the diminishing marginal product of capital. As capital accumulates relative to labor, the effect of further capital accumulation on output and wages gets smaller. Formally, note that the effect of an increase in k on the marginal product of labor is $(1-\alpha)\alpha A k^{\alpha-1} = (1-\alpha)\times$*marginal product of capital.*

and the number of workers remains the same, the effect on production will be the same as doubling the workforce. We further assume that technical progress is such that D_t increases from one period to the next at the constant rate, d. Thus, $H_{t+1}/H_t = n(1 + d)$, the effective workforce increases due to both population growth and technical progress.

We can model the firms as choosing H_t and paying a wage rate per unit of *effective* labor, w_t, a slight change from the previous interpretation. The total wage payment received by an actual worker will now be w_tD_t. The factor price equations given by (2.2a, 2.2b) remain the same, except we now must interpret k as the capital to effective labor ratio, i.e. $k = K/H$.

Now let's think about how the equilibrium and transition equation are altered by technical progress. The firm's demand for the capital, which we can think of as a demand for the ratio of capital to effective-labor, will take the same form as (2.2a). On the household side, we need only adjust the saving function for the new concept of household wages to get $s_t = \frac{\beta}{1+\beta} w_t D_t$. The supply of capital per effective worker is defined as $k_t^s \equiv s_{t-1}N_{t-1}/D_tN_t$. Using the household saving function, the supply of capital per effective worker can be written as $k_t^s = \left[\frac{\beta}{1+\beta}\right]\frac{w_{t-1}}{n(1+d)}$. Finally, using the factor price equations, the adjusted transition equation becomes

$$k_t = \left[\frac{\beta}{1+\beta}\right]\frac{(1-\alpha)Ak_{t-1}^{\alpha}}{n(1+d)}, \tag{2.9}$$

which has the same form as (2.8), except for the presence of $1 + d$ in the denominator of the expression on the right-hand-side of the equation.

Thus, the transitional dynamics of the economy are the same as before. However, now there is an *endogenous* source of growth (increasing physical capital intensity) and an *exogenous* source of growth (technical progress). When the steady state is reached, the transitional growth from increasing physical capital intensity is over and interest rates become constant. However, there will continue to be positive economic growth from exogenous technical progress. Labor productivity, real wages per worker (w_tD_t), and the standard of living (measured by consumption per household), all increase at the rate $d > 0$ in the steady state.

2.4 Quantitative Theory

Over the last 40 years there has been an increasing tendency for macroeconomists to *quantify* their theoretical models. Quantifying a model means determining numerical values for the model's parameters, thereby enabling the model to generate numerical predictions that can be compared to real world data. This healthy tendency to develop theories that can be quantified has greatly improved the understanding of many different phenomena and has created a progressive scientific paradigm within which to conduct macroeconomic research.

In this section, we quantify our simple growth model and compare its predictions to important qualitative patterns we commonly see in the data as economies grow. We are effectively repeating a version of the exercise conducted in the famous article by King and Rebelo (1993). They showed that the standard neoclassical model of physical capital accumulation is not consistent with the pattern of growth rates and interest rates experienced by the U.S. as it developed.

In most cases it is not possible to use the traditional econometric approach of parameter estimation (due to a desire to limit the number of variables in the analysis, the nonlinear structure of the model, or the lack of appropriate data). Instead the model is *calibrated*. That is parameters are set so as to allow the model to *match* certain *targets*—observations or previously estimated behavioral responses.[5] Once calibrated, the model can generate predictions about the values of variables that were *not* used in the calibration. The predicted values can then be compared against data to assess the model's ability to replicate the real world. Failures to replicate important real world observations then lead an adjusted model, or an entirely new model, that provides a better approximation. The model currently providing the best approximation should be favored to conduct *policy analysis*, where the effects of current and proposed government policies are evaluated. Continually pursuing the most accurate quantitative approximation is the best chance we have of improving our understanding of economies and policies.

Let's make these ideas more concrete by calibrating a simple neoclassical model of physical capital accumulation and then testing its predictions about economic growth. The transition Eq. (2.9) provides the basic model. The equation contains six exogenous parameters: α, β, δ d, A, and n. To allow for endogenous growth through increasing physical capital intensity, we will have to start the economy in an initial position that is below its steady state. So, an initial value, k_1, will also have to be determined. Finally, the length of each time period in the model must be chosen. In fact, some parameter values will depend on the time-period choice.

Part of the calibration typically involves matching the steady state of the model to certain observations (for example, the interest rate or return to capital). Since all variables in the neoclassical growth model can be related back to k, we will need the steady state solution of (2.9),

[5]There are differences of opinion about what qualifies as an appropriate target. Some believe that calibration should *not* involve previous econometric estimation. According to this view, all parameters within a model should be set to match particular data points or statistical moments of a data set (sample means, variances, and covariances), but *not* to match econometric estimates found in the literature. Others broaden the targets to include previous statistical estimates of the model's parameters and behavioral responses, *even if the model used in the estimation is not the same as the one used in the calibration*. We are comfortable with either approach. The important point from our perspective is that all quantitative models, *however calibrated*, should be tested by comparing their predictions against observations or statistics *not* used in the calibration process. The fact that these "tests" or comparisons are not as formal and refined as traditional hypothesis testing in statistics does not particularly concern us. At this stage in the profession's understanding of macroeconomics, models that even roughly approximate reality are difficult to find. Hopefully, as our approximations become more refined, we will need to worry about more formal testing procedure.

$$\bar{k} = \left[\frac{\beta A}{1+\beta}\frac{1-\alpha}{n(1+d)}\right]^{\frac{1}{1-\alpha}}. \tag{2.10}$$

2.4.1 Calibration

In our two-period life-cycle model, the first period is designated the "work-period" and the second period the "retirement –period." In this setting, it is often assumed that each period lasts 30 years. In comparison to the real world, a 30-year period makes the working life too short and the retirement period too long. The more periods we allow in the life-cycle, the more realistic the model becomes.

For example, we could instead assume that *three* twenty-year periods represent a lifetime, with two working periods (40 years) and one retirement period (20 years). However, as you add periods, the model becomes more complicated. Each additional period of life added, also adds a new generation to the economy. In a life-cycle model where each household lives for three periods, there will be a young, a middle-aged, and an old household alive in any given time period. The complication of keeping track of different generations is a clear disadvantage of using an explicit overlapping generations approach. However, advances in computing are lessening the disadvantage over time. In this book, we stick with a two-period model because it is sufficient to generate several important qualitative and quantitative implications.[6]

With the time period selected, we can begin setting other parameter values. Our application will examine the model's ability to explain growth in the U.S. from the end of the Civil War through the end of the twentieth century. In applying the model, a useful way to proceed is to create a relatively simple *baseline* calibration and then do a *sensitivity analysis* by examining how results change as we deviate from the baseline calibration or model specification.

The annual rate of population growth actually fell over this historical period, from 2.3% in the late nineteenth century to about 1% by the end of the twentieth century (Barro 1997). For the baseline calibration we simplify and set the annual rate of population growth to be 1% over all periods. Time periods in the model last 30 years, so the value for population growth in the model is the one percentage point annual rate of growth compounded for 30 years. The value of n is then chosen to satisfy the equation $n = (1.01)^{30} = 1.3478$.

The capital share of output and income has shown no systematic trend in U.S. history or across countries at different stages of development today (Gollin 2002). We set α to a commonly estimated value of 1/3. The annual rate of depreciation on physical capital is estimated to be in a range between 5 and 10% (e.g. Stokey and Rebelo 1995). We set the annual rate of depreciation to 7%. To translate the annual depreciation rate into the depreciation rate over 30 years, think about how

[6] For a further discussion of the issues associated with quantifying overlapping generations models see Appendix B.

much capital remains each year after depreciation occurs. In any given year the physical capital stock at year's end is 93% of its value at the beginning of the year. If you start with one unit of capital today, then after 30 years there would be $1 - \delta = (1 - 0.07)^{30} = 0.93^{30} = 0.1134$ units of capital. So, $\delta = 0.8866$.

Note that we can write worker productivity or output per worker as

$$\frac{Y_t}{M_t} = \frac{Ak_t^\alpha D_t M_t}{M_t} = Ak_t^\alpha D_t. \tag{2.11}$$

So we can write the ratio of worker productivity in 1990 to worker productivity in 1870 as

$$\frac{(Y/M)_{1990}}{(Y/M)_{1870}} = \left(\frac{k_{1990}}{k_{1870}}\right)^\alpha \frac{D_{1990}}{D_{1870}}. \tag{2.12}$$

For the baseline case, we arbitrarily set d so that exogenous technical progress explains "half" the economy's growth. The annual rate of growth in labor productivity from 1870 to 1990 was about 1.6% (Rangazas 2002). With a growth rate of 1.6% per year over 120 years, labor productivity was 6.7180 times higher in 1990 than in 1870. In terms of a geometric mean, half of this growth is $6.7180^{1/2} = 2.5919$. The annual rate of technical progress needed to generate this much growth is 0.7968%. This means that $1 + d = (1.007968)^{30} = 1.2688$, or $d = 0.2688$.

Finally, we set β to match the rate of return to capital. We take the rate of return to capital to be the rate of return on the Standard and Poor's 500 over the twentieth century. The annual real rate of return on this portfolio of stocks averaged 7% over the twentieth century (Kocherlakota 1996). Due the absence of any trend in the annual rate of return over the century, we assume that the U.S. economy was close to its steady state at least by the end of the twentieth century. Thus, we have $1 + \bar{r} - \delta = 1.07^{30} = 7.6123$. Using (2.2a) and (2.10), we have

$$\frac{\beta}{1+\beta} = \frac{\frac{\alpha}{1-\alpha} n(1+d)}{\bar{r}}. \tag{2.13}$$

Plugging the calibrated values of the other parameters into (2.13) implies $\beta = 0.1287$.

We still have to set the initial value of k_t. The idea is to set k_1 so that half of the economy's growth is explained by capital accumulation (that portion not explained by technical progress). So, choose k_1 to satisfy

$$\left(\frac{k_5}{k_1}\right)^\alpha = 2.5919. \tag{2.14}$$

Since both k values in (2.14) are unknown before the model simulation is run, we have to experiment with values for k_1 until we find one that satisfies (2.14).

By assuming that the economy is close to its steady state in 1990, we can get a good guess for k_1 by using (2.14) to write $k_1 = \bar{k}/17.41$. To determine the absolute values of $\bar{k}$ and k_1, we need to set a value for A. This parameter is different than the others because it only scales the level of production. There is no particular reason for us to replicate the *level* of production observed in the real world (even the real-world index numbers for GDP are arbitrary). So, we set A to be one. This implies $\bar{k} = 0.00937$ and, as an initial guess, $k_1 = 0.000538$.

To summarize, the calibrated parameters are given below.

Calibration

n	1.3478
d	0.2688
A	1.0000
α	0.3333
β	0.1287
δ	0.8866

2.4.2 Historical Simulation

We are now ready to do a historical simulation. Plug the guess for k_1 into (2.9) and let the model generate values for k_t. Change the guess for k_1 until (2.14) is met. Once finding values for k_t that satisfy (2.14), then compute the predicted interest rates and labor productivity growth rates. The annualized values of predicted interest rates (solid line) and growth rates (dashed line) are displayed in Fig. 2.3.

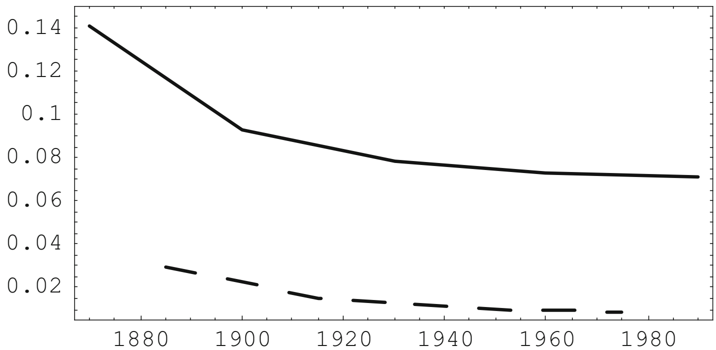


Fig. 2.3 Simulated U.S. interest rates and growth rates: 1870–2000
Notes: The solid line gives the annualized rate of return to capital and the dashed line gives the annualized growth rate of labor productivity. The annualized growth rates over 30 year periods were plotted above the midpoint of the intervals between the periods

The model predicts high interest rates (14%) and growth rates (3%) for the late nineteenth century and then a decline in both variables over the twentieth century. These predictions miss the mark for a number of reasons.

Returns to capital were probably higher in the late nineteenth century than during the twentieth century. We do not have returns on the Standard and Poor's 500 that go back as far as 1870, but the returns on other assets were 2–6 percentage points higher in 1870 than in the twentieth century. Wallis (2000, Fig. 2.2) reports that real interest rates on national government debt averaged about 5% in the first half of the nineteenth century and averaged about 2.5% during the twentieth century. Barro (1997) reports that real interest rates on commercial paper were 9% from 1840 to 1880, but averaged about 3% during the twentieth century. The model predicts initial interest rates out of this range, about 7 percentage points higher than in the twentieth century. Also by 1900, interest rates showed no trend, while the model predicts a downward trend throughout the twentieth century, especially in the first third of the century.

The growth rate predictions are even less accurate. Table 2.1 presents estimates of U.S. labor productivity growth rates for two centuries (Mourmouras and Rangazas 2009). Growth rates showed little trend from 1840 to 2000. In contrast, the model predicts high growth rates in the nineteenth century and then a steady decline.

The fundamental problem with the standard neoclassical growth model is clear. In order to satisfy (2.14), the capital-labor ratio must be set well below its steady state value in 1870. The relatively low capital-labor ratio produces relatively high returns to capital. The fact that the capital-labor ratio is well below its steady state value generates high and declining growth rates, as indicated qualitatively by the transition equation diagram in Fig. 2.2.

The only way to make the model's predictions more accurate is to set k_1 closer to $\bar{k}$. But this means much less than half the historical growth will be explained by physical capital accumulation. More endogenous sources of growth are needed to produce a satisfactory explanation of growth in United States history. One of these sources is public capital in the form of public schooling, roads, public utilities, and

Table 2.1 Growth rate in output per worker

1820	0.31
1840	1.82
1860	1.32
1880	1.84
1900	1.53
1920	1.40
1940	1.72
1960	2.45
1980	1.58
2000	1.62

Notes: The Table gives annual growth rates in worker productivity over two centuries of U.S. history. See Mourmouras and Rangazas (2009) for sources other aspects of government infrastructure. The next section extends the model to include public schooling, an investment in human capital. Chapter 3 adds public capital more generally.

2.5 Human Capital

One reaction to failure of physical capital accumulation to explain much growth has been to introduce human capital. Human capital is the knowledge and skill embodied in a worker that increases productivity. Human capital investments increased dramatically over the twentieth century in the United States and many economists view human capital as an important source of economic growth that is ignored in the standard model.

While human capital can be formed through a variety of different investment activities, the primary focus has been on formal education received in schools. This focus is largely because of data availability. Learning away from school and on-the-job training are likely important but harder to measure. Investments in worker's health should also be included but typically are not.

Table 2.2 presents two measures of formal education investments in children: real spending per child in primary and secondary school (x_t) and the fraction of the year spent in school by children ages 0–19 years (e_t). The average time spent in school by an age cohort rises because children attend school for more years and because there is a rise in the days attending school per year. School spending per pupil in school expanded more than 25-fold since 1870 and time spent in school expanded more than three-fold.

To identify the growth implications of increasing education investments we use a human capital production function. We extend our effective labor input definition to now include the impact of education on worker productivity,

$$H_t = D_t^{1-\theta_1} x_{t-1}^{\theta_1} e_{t-1}^{\theta_2} M_t, \tag{2.15}$$

where θ_1 and θ_2 are constant parameters that capture the effect of school spending and student time on human capital. We assume the parameters satisfy the restriction $0 \leq \theta_1 + \theta_2 \leq 1$. Note that the model of exogenous effective labor supply from Sect. 2.3 can be obtained by setting $\theta_1 = \theta_2 = 0$.

Table 2.2 Human capital investments in the U.S.

Investment	1840	1870	1900	1930	1960	1990
x_t	1.0	1.0	1.8	4.6	9.5	25.2
e_t	0.08	0.09	0.11	0.215	0.29	0.30

Notes: School spending is defined as total expenditures per pupil. The first row gives the real spending per pupil for each year divided by the real spending for the 1870 (Rangazas 2002). The time investment data are from Lord and Rangazas (2006). The expenditure value for 1840 is assumed the same as for 1870 because there was little change in spending between 1870 and 1880. The student time value for 1840 is assumed to be the same as for 1850

To get an initial estimate of the potential importance of education in explaining historical growth in the U.S. we take the flowing steps.

1. Use the data from Table 2.2 as exogenous education investments. Later in the Chapter and in Chap. 4 we present theories of human capital investment that can be used to attempt an explanation of the education data, but one step at a time.[7]
2. Find reasonable estimates of θ_1 and θ_2.
3. Use (2.15) as a substitute for the previous notion of H_t and adjust the growth model used in Sect. 2.3 accordingly.
4. Re-do the historical simulation from Sect. 2.4.

Based on econometric evidence, and consistency with empirical estimates of the rates of return to human capital investments, reasonable settings for the human capital production function parameters are $\theta_1 = 0.10$ and $\theta_2 = 0.40$ (see a complete discussion in Rangazas 2002). We de-trend the school spending data by writing (2.15) as $H_t = D_t M_t h_t$, where $h_t = \tilde{x}_{t-1}^{\theta_1} e_{t-1}^{\theta_2}$, with $\tilde{x}_t \equiv x_t / D_t$. The reason for writing (2.15) this way is to generate a conservative estimate of the role of education spending. A big part of education spending is paying teachers. Teachers' pay rises when general wages rise in the economy as a whole in order to prevent them from quitting and working elsewhere. Technological progress is believed to raise worker productivity and wages primarily in non-teaching occupations. So, technological progress *increases the cost* of paying teachers but *not their productivity* as teachers (this is known as "Baumol's Disease"). The de-trended measure eliminates this rise in cost and better identifies the spending associated with an increase in education inputs, e.g. smaller class sizes (more teachers) and more books and computers.

Next, treat the new interpretation of H_{t+1} just as $D_{t+1}M_{t+1}$ was treated in deriving the transition equation at the end of Sect. 2.3. Using the same algebraic approach followed previously, the adjusted transition equation with schooling is

$$k_t = \left[\frac{\beta}{1+\beta}\right] \frac{(1-\alpha)Ak_{t-1}^{\alpha}}{n(1+d)} \frac{h_{t-1}}{h_t}, \tag{2.16}$$

where $k \equiv K/H$. Note, because physical capital intensity is measured relative to the value of H, the growth in effective labor supply due to education lowers k in the same way as d.

The simulation experiment is conducted in the same manner as in the model without human capital. Let period 0 denote 1840. The initial value of k is determined by

[7]One can think of the financing of public school expenditures as coming from a tax on capital or capital income (similar to the property tax used to finance schooling in the U.S.). When $\sigma = 1$, as with our log preferences, a capital tax has no effect on saving and the transition equation.

$$\left(\frac{k_5}{k_1}\right)^{\alpha}\frac{h_5}{h_1} = \left(\frac{k_5}{k_1}\right)^{\alpha}\left(\frac{\tilde{x}_4}{\tilde{x}_0}\right)^{\theta_1}\left(\frac{e_4}{e_0}\right)^{\theta_2} = 2.5919 \tag{2.17}$$

Similar to Eqs. (2.14), (2.17) requires that physical and human capital accounts for "half" the growth from 1870 to 1990, with the other half accounted for by exogenous technical progress. Physical capital intensity is now getting help from human capital, so k_1 will not have to be as far below k_5 in order to satisfy (2.17) as was necessary when using (2.14) in our first simulation. This feature should help in eliminating the counterfactual predictions given in Fig. 2.3. Finally, note that while human capital investments rose sharply over the period, both education inputs are subject to diminishing returns because θ_1 and θ_2 are both significantly less than one. Thus, there will be two opposing forces on the economy's growth rates—rising rates of human capital investment versus diminishing returns to those investments.

Figure 2.4 presents the model's historical predictions regarding interest rates and growth rates. The plots of both time series have flattened considerably compared to the series in Fig. 2.3. Growth rates now show no trend over the entire period and interest rates no trend over the twentieth century—making both time series more consistent with the data. Capital accumulation, physical and human, can explain a 2.6-fold increase in worker productivity over this historical period while maintaining consistency with the time paths of interest rates and growth rates.

The slow growth in human capital inputs caused human capital growth to be slow before 1900. As a result, much of the model economy's growth before 1900 was due to rising physical capital intensity. This is clear from the approximately 2 percentage point drop in interest rates from 1870 to 1900—within the range of the decline seen in the historical interest rate data. The late ninetieth century was when the Industrial Revolution intensified in the United States. During the twentieth century, physical capital intensity and interest rates showed no trend—in the model and in the data. The explained growth over this period was due to rising human capital investments.

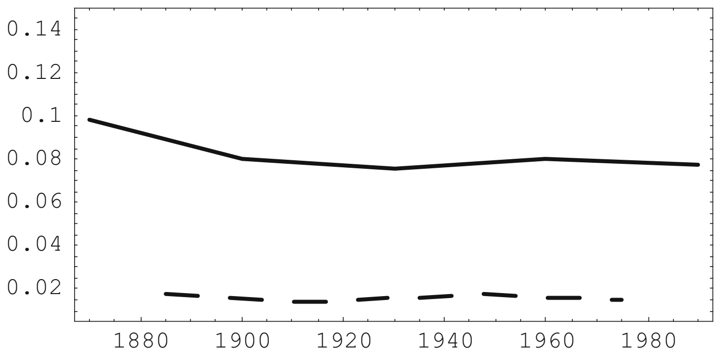


Fig. 2.4 Simulated U.S. interest rates and growth rates: 1870–2000—with human capital
Notes: The solid line gives the annualized rate of return to capital and the dashed line gives the annualized growth rate of labor productivity. The annualized growth rates over 30 year periods were plotted above the midpoint of the intervals between the periods

However, worker productivity growth rates in the model showed no trend during the twentieth century, also consistent with the data, because the diminishing returns to human capital investments roughly offset the effect of the rising human capital investments. Overall, the inclusion of human capital significantly improves the fit to the historical data.

2.6 Intergenerational Transfers

The analysis of Sect. 2.5 did not include a theory of why human capital investment rose over time to increase worker productivity. This section uses theories of *intergenerational transfers* that attempt to explain human and physical investments by parents designed to increase the wealth of their children. In his presidential address to the American Economic Association, Gary Becker (1988) encouraged economists to pay greater attention to the role of the family in their thinking about macroeconomic issues. Becker pointed out how intergenerational transfers between family members are likely to influence economic growth and alter the effects of important fiscal policies such as social security and government borrowing. Understanding the causes and consequences of intergenerational transfers is one of the most important motivations for taking an overlapping generations approach that recognizes the generational structure of the economy.

We define intergenerational transfers to be a private transfer of resources from one generation to another generation. In this section the transfers take two distinct forms: *human capital* investments (x) and transfers of physical or financial assets (b). The assets are identical to the assets used for life-cycle saving, but instead of being used to increase future income for the household, they are used to increase the future income of the household's children.

Human capital investments are *in-kind* transfers of goods and services designed to raise the recipient's productivity by increasing knowledge, skills, or health. The effect of these in-kind transfers on the child's market productivity is given by a human capital production function $h_{t+1} = h(x_t)$. We temporarily ignore time investments by the children themselves. Time investments can be incorporated into the framework because parents largely decide how much time their children spend in school and how much time they spend working to bring income to the family. Time investment will be added in Chap. 4. With the focus on the endogenous human capital component of effective labor supply, we also temporarily ignore the exogenous component of effective labor supply, D_t.

The only assumption about the human capital production function that we will use in this section is that the derivative of h with respect to x is positive but diminishing ($h' > 0$, $h'' < 0$), i.e. h is increasing and strictly concave in x. This assumption means that while spending more on children's education this period will always increase their knowledge and productivity next period, when they enter the labor force, the marginal return to additional educational spending decreases with the level of expenditures. In other words there are natural limits to the rate at which a child can accumulate skills and knowledge.

The distinction between the two types of transfers is important since there are situations where human and financial investments of equal dollar amounts are not equivalent in their effect on the recipient's behavior and wealth. For example, a dollar spent on a dependent child's primary and secondary education is generally not going to have the same economic effect as a dollar invested in a financial asset that is ultimately bequeathed to the child. In other circumstances the distinction may be less clear. Is the payment of a child's college tuition by a parent equivalent to a cash transfer to the child? The answer is not obvious and will depend on the preferences and constraints of *both* the parent and the child. It may also be true that parents value the two transfers differently. For example, parents may value education for its own sake, beyond its effect on the child's future labor productivity in the market.

The accounting relationship between intergenerational transfers and the accumulation of wealth is made precise by examining the household budget constraints in our two-period model. Households have n children, an exogenous variable in this chapter, in the first period of adulthood. During the first period, along with the typical life cycle consumption and saving choices, parents make human capital investments in their children and save to make financial transfers to their children when they become young adults in the next period. The first period budget constraint is

$$c_{1t} + s_t + n(x_t + b_{t+1}) = W_t, \tag{2.18}$$

where $W_t \equiv w_t h_t + R_{t-1} b_t$ is initial wealth and recall that $w_t h_t$ now represents lifetime earnings with w_t interpreted as the wage per unit of *effective* labor supply.[8] The other term on the right-hand-side of the initial wealth expression is the physical or financial asset, plus interest, that the household receives from its parents during the first period of adulthood. The second period constraint remains the same as before and given by

$$c_{2t+1} = R_1 s_t. \tag{2.19}$$

In addition to the budget constraints, one important additional constraint is needed. It is assumed that parents cannot accumulate debt that their children are legally bound to pay (a legal restriction that exists in most societies). This means there is a nonnegativity constraint on financial transfers

$$b_{t+1} \geq 0. \tag{2.20}$$

[8]Note that the acquisition of financial assets occurs in the first period of adulthood and the financial transfers are made in the second period of adulthood. Thus, the transfers are made when both generations are alive. Transfers of this type are called *inter vivos* transfers, as opposed to *bequests* that are transferred at death. In our model, where we assume (i) perfect certainty, (ii) perfect life-cycle credit markets, and (iii) no strategic interactions between generations, the timing of financial transfers is irrelevant. However, the timing of transfers can matter when these conditions are not met (see for example, Bernheim et al. (1985) and Cox 1987).

We next turn to two different approaches to modeling *why* parents make intergenerational transfers. The first approach assumes parental altruism, parents care about the welfare of their children. The altruistic approach provides an interesting and rich theory, but also one that is technically demanding. Less advanced students should skip to the second theory of transfers. This theory assumes that parents receive utility, a "warm glow", directly from the transfers they make to their children.

2.6.1 Altruism

One natural way to explain *why* intergenerational transfers occur is to assume that parents are altruistic, i.e. that they care about their children's welfare as adults. An important way of modeling altruism assumes that parents care about their adult children's lifetime utility. We call this type of altruism *Barro-Becker* altruism named after the two economists who introduced the concept (Barro 1974; Becker 1974). As we shall see, this particular way of modeling altruism has very interesting implications, ones that have had a major effect on the development of macroeconomic theory.

Barro-Becker altruism implies a utility function for the generation-t parent of the form, $U_t + \beta V_{t+1}(W_{t+1})$, where U_t is utility from the consumption of the generation-t household, which we take to be of CES form developed in *Problem* 7 throughout our discussion, and V is the *maximum attainable utility* of the next generation. The function V_{t+1} depends on W_{t+1} because it's the utility attained when adult children maximize their utility subject to their initial wealth and market prices.[9] We assume that parents cannot directly affect the choices of their adult children, but they can have a significant effect on their children's initial wealth and therefore on V. The value function is discounted by parents because the adult utility of their children is generated in the future and the parents have a positive rate of time preference. It is also possible that parents care more or less about their adult children's welfare than their own. In this case, the generational discount factor would differ from the pure time discount factor, but the qualitative conclusions would be the same.

[9]In microeconomics, the maximum attainable utility function is called an *indirect utility function*. In the pure life cycle version of our model, with no altruism, an indirect utility function is easily obtained. Take the optimal consumption choices of the household (5) and substitute them back into the CES utility function. For example, if we do this for the case of $\sigma = 1$, we get the very simple indirect utility function $U_t^* = \beta \ln \beta + (1+\beta) \ln \frac{1}{1+\beta} + (1+\beta) \ln w_t + \beta \ln R_t$. With altruism, things are not nearly so simple. This is because generation-t's utility depends on generation-$t+1$'s utility, which depends on generation-$t+2$'s utility, and so on. As we shall see, in this case V_t cannot be found directly. Instead it is implicitly defined by a difference equation in the function V_t— what is known as a *Bellman equation*. In this case, V_t is called a *value function* and solving for this function is tricky business. Stokey and Lucas (1989) provide a general discussion of the conditions under which the value function exists, is unique, and is differentiable with respect to initial wealth.

We are now ready to formally lay out the optimization problem for the altruistic household. The generation-t altruistic household chooses life-cycle consumption and intergenerational transfers to maximize

$$U_t + \beta V_{t+1}(W_{t+1}), \tag{2.21}$$

subject to $W_{t+1} \equiv w_{t+1}h_{t+1} + R_t b_{t+1}$ and the generation-t lifetime budget constraint, formed by combining (2.18) and (2.19), along with (2.20). The problem can be solved in two steps. First, for *given* values of the transfer variables, the solution for life-cycle consumption essentially follows as before

$$c_{1t} = \Psi_{1t}(W_t - n(x_t + b_{t+1})) \tag{2.22a}$$

$$c_{2t+1} = [\beta R_t]^{\sigma} c_{1t}, \tag{2.22b}$$

Where wealth available for generation-t consumption is now reduced by transfers made to their children. The life-cycle consumption choices can then be substituted back into (2.21), making it solely a function of the intergenerational transfer choices.

Second, (2.21) can be maximized with respect to the transfer choices. The first order conditions for the transfer choices are

$$n(W_t - n(x_t + b_{t+1}))^{-1/\sigma}\Psi_{1t}^{-1/\sigma} = \beta V'_{t+1}h'_{t+1} \tag{2.23a}$$

$$n(W_t - n(x_t + b_{t+1}))^{-1/\sigma}\Psi_{1t}^{-1/\sigma} \geq \beta V'_{t+1}R_t \tag{2.23b}$$

The left-hand-side of each equation is the marginal cost of making a transfer to children measured in terms of forgone consumption and utility of the parent. The right-hand-side is the marginal benefit of the transfer measured in terms of the resulting rise in lifetime wealth and utility of the adult child. The inequality in (2.23b) includes the possibility that a nonnegative financial transfer may not be optimal. If not, (2.20) binds and (2.23b) becomes a strict inequality with the interpretation that the marginal cost of the financial transfer at $b_{t+1} = 0$ is greater than the marginal benefit.

Note that solving this maximization problem gives us the maximum attainable utility, or *value function*, of the current generation,

$$V_t(W_t) = \max\{U_t + \beta V_{t+1}(W_{t+1})\}.$$

This equation implicitly defines the value function and is known as the *Bellman equation*. The Bellman equation allows us to obtain the following result (see Appendix A)

$$V'_{t+1} = (W_{t+1} - n(x_{t+1} + b_{t+2}))^{-1/\sigma}\Psi_{t+1}^{-1/\sigma} = c_{1t+1}^{-1/\sigma}, \tag{2.24}$$

The derivative of the value function is just the marginal utility of consumption. Using (2.24), the first order conditions given by (2.23a, 2.23b) can be presented and discussed intuitively for the two possible cases,

unconstrained transfer equations

$$w_{t+1}h'(x_t) = R_t \tag{2.25a}$$

$$c_{1t+1} = [\beta R_t/n]^{\sigma} c_{1t} \tag{2.25b}$$

or

constrained transfer equations

$$b_{t+1} = 0, \tag{2.26a}$$

$$c_{1t+1} = [\beta w_{t+1}h'(x_t)/n]^{\sigma} c_{1t}. \tag{2.26b}$$

If the optimal financial transfers are positive, then human capital investments are unconstrained. This allows the investments to be *productively efficient* in the sense that the return on human investment is equated to the return on physical investment. Parents invest in their children until the rate of return to human capital is driven down to the rate of return on physical capital. At that point any further intergenerational transfers are in the form of physical or financial assets.

In the efficient case, human capital investment is completely determined by (2.25a), independent of consumption choices and household wealth. In addition, when the generations are linked by financial transfers, one gets the familiar Euler condition, (2.25b), where the growth of consumption over time (actually over generations in this case) is determined by the return to capital—perfectly analogous to the optimal pattern of life-cycle consumption growth in (2.22b).

Desired financial transfers may also be negative, i.e. parents may want to leave debt for their children to repay. In this case (2.20) binds and there is a strict inequality in (2.23b). Now the human capital investment decision cannot be separated from the household's consumption choices; (2.22a, 2.22b) and (2.26b) must be solved simultaneously. Under these circumstances, the *wealth* of the current generation becomes a determinant of human capital investment. Combining (2.23a) and (2.23b) shows that this situation is *productively inefficient*, because the rate of return on human capital investment exceeds the rate of return on financial assets. However, the current generation cannot "afford" to invest more in their children's education, which is why they want to raise both their own consumption and human capital investment in their children by leaving debt for their children to pay. The legal restrictions forbids this, so they are forced to set $b_{t+1} = 0$. If parents could invest more, by setting $b_{t+1} = 0$, then x_t would rise and h' would fall until (2.25a) is satisfied. This situation implies that the ratio of consumption of the next generation to the consumption of the current generation would also fall.

2.6.2 Explicit Household Level Solutions in Some Special Cases

To go further in uncovering the behavioral implications of adding altruistic transfers we now solve the model for some special cases. For most of the special cases, we also assume that the economy is in a steady state with zero technical progress, so that $r_t = r$ and $w_t = w$. This assumption serves to simplify the notation considerably.

2.6.2.1 Exogenous Human Capital

To isolate the role of financial transfers, assume that human capital is exogenous and constant, and, for simplicity, set $h = 1$. Under this assumption, a household can only help its descendants by giving them financial transfers. In this case the *constrained* solution, with $b_{t+1} > 0$ for all t, reverts back to the pure life-cycle solution—with no links between the generations. To solve the *unconstrained* case, with $b_{t+1} > 0$ for all t, we begin by writing out the lifetime budget constraints for the first two generations

$$c_{1t} + \frac{c_{2t+1}}{R} + nb_{t+1} = w + Rb_t \tag{2.27a}$$

$$c_{1t+1} + \frac{c_{2t+2}}{R} + nb_{t+2} = w + Rb_{t+1}. \tag{2.27b}$$

Solving for b_{t+1} in (2.27a), substituting into (2.27b), and rearranging gives

$$c_{1t} + \frac{c_{2t+1}}{R} + \left(\frac{n}{R}\right)\left[c_{1t+1} + \frac{c_{2t+2}}{R}\right] + \left(\frac{n}{R}\right)nb_{t+2} = Rb_t + w + \left(\frac{n}{R}\right)w. \tag{2.28}$$

Notice that we could solve (2.28) for b_{t+2} and substitute the solution into the generation-$t + 2$ version of (2.27a, 2.27b) to get a version of (2.28) for *three* generations of the family.

Proceeding in this way, by successively substituting into the lifetime budget constraints of future generations, produces a budget constraint for the entire "family dynasty,"

$$c_{1t} + \frac{c_{2t+1}}{R} + \left(\frac{n}{R}\right)\left[c_{1t+1} + \frac{c_{2t+2}}{R}\right] + \left(\frac{n}{R}\right)^2\left[c_{1t+2} + \frac{c_{2t+3}}{R}\right] + \cdots = W_\infty, \tag{2.29}$$

where $W_\infty \equiv Rb_t + w + \left(\frac{n}{R}\right)w + \left(\frac{n}{R}\right)^2 w + \cdots$. Thus, W_∞ is the wealth of the entire family dynasty, a well-defined value provided that rate of return on assets exceeds the sum of population growth rate and the growth rate in wages (zero in this case) over time.

The left hand side of (2.29) can be simplified by using the Euler equation for life-cycle consumption, (2.22b), to get

$$\Psi_1^{-1}c_{1t} + \left(\frac{n}{R}\right)\left[\Psi_1^{-1}c_{1t+1}\right] + \left(\frac{n}{R}\right)^2\left[\Psi_1^{-1}c_{1t+2}\right] + \cdots = W_\infty. \tag{2.30}$$

Next, use (2.25b), the Euler equation that applies to generations, to express consumption for *each* generation in terms of consumption for the *first* generation to get

$$\Psi_1^{-1} c_{1t}\left\{1 + \left(\frac{n}{R}\right)\left[\frac{\beta R}{n}\right]^{\sigma} + \left(\frac{n}{R}\right)^2\left[\frac{\beta R}{n}\right]^{2\sigma} + \ldots\right\} = W_{\infty}. \tag{2.31}$$

The geometric sum in the curly brackets is finite provided $\beta^{\sigma}\left(\frac{n}{R}\right)^{1-\sigma} < 1$ or $\beta^{\sigma/(1-\sigma)} < \frac{R}{n}$. If $R > n$, then this condition holds if $\sigma \leq 1$ and $\beta \leq 1$ Under these conditions, Eq. (2.22a, 2.22b) then gives us the solution

$$c_{1t} = \Psi_1\left(1 - \beta^{\sigma}\left(\frac{n}{R}\right)^{1-\sigma}\right)W_{\infty}, \tag{2.32}$$

which says the consumption of the current generation is a function of its wealth *and* the wealth of all future generations as well. Once the current generation makes its consumption choice, then by using (2.25b) we can find the consumption of *every* member of the family dynasty. Thus, the current generation determines the family consumption path into the indefinite future. In this sense, the current generation behaves "as if" it is "infinitely-lived".

The *infinitely-lived agent* model is perhaps the dominant model for studying macroeconomics because it allows one to avoid treating each generation as a distinct household. Being able to ignore the generational structure of the economy is a very handy simplification, especially when you have many generations co-existing in the same period. In *Problem* 21, you will see that the infinitely-lived model allows the entire economy's behavior to be determined by a single *representative agent*.

However, the assumptions that must hold for the infinitely-lived agent to be valid are strong. In particular, intergenerational transfers must be strictly motivated by Barro-Becker altruism and financial transfers must be nonzero for all generations. Evidence presented in Sect. 2.7 suggests that these assumptions do not generally hold. Yet, for some purposes, the convenience of the infinitely-lived agent model may be useful without doing obvious harm to the analysis. For example, to study business cycles each period must represent a year or even a quarter of a year. Thus, you cannot aggregate time as we have in the two-period life-cycle model. If each period corresponds to a year, then the life cycle of an adult must contain 50–60 periods. This, in turn, means that *many* generations will be present in *each* period of the model. It simplifies the analysis greatly if this generational structure can be ignored. In addition, business cycles do not cause resources to be shifted across generations in any obvious systematic fashion. Here, the argument for the infinitely lived agent model is convincing. For other issues, in particular long-run issues that involve significant transfers of resources across generations, use of the infinitely-lived agent model can be very misleading.

It is possible to derive an explicit solution for the optimal financial transfers. This solution adds additional insights and gives a sense of the dramatic implications of assuming intergenerational altruism. Using (2.22b), (2.27a), and (2.32) we get

$$nb_{t+1} = W_t - \left(1 - \beta^{\alpha}\left[\frac{n}{R}\right]^{1-\sigma}\right)W_{\infty}. \tag{2.33}$$

Holding W_{∞} constant, an increase in W_t will raise transfers to children one for one. Note if W_{∞} is constant and W_t increases, then the wealth of future generations must have fallen in present value by the amount that W_t rose (so as to maintain W_{∞}). In this thought-experiment, the rise in the wealth of the current generation is entirely at the expense of future generations. The current generation acts to exactly offset the exogenous reallocation of the dynasty's wealth by transferring the entire increase in W_t back to future generations. It does this by increasing transfers to its children. Thus, any exogenous reallocation of dynasty wealth is completely undone by endogenous intergenerational transfers in the opposite direction.

Next consider a rise in W_{∞}, holding W_t constant. In this thought experiment, intergenerational transfers fall. To see the intuition, if W_{∞} increases while holding W_t constant, then one or more future generations in the dynasty must have experienced a rise in wealth. The only way that the current generation can share in the rise in dynastic wealth is by reducing transfers to the future generation, allowing its consumption to rise as indicated by (2.32).

Finally, consider a rise in W_t with no change in the wealth of future generations. In this case, W_{∞} rises by the same amount as the rise in W_t. The change in transfers is $\beta^{\sigma}\left(\frac{n}{R}\right)^{1-\sigma} < 1$ times the change in W_t. Thus, part of the rise in the wealth of the current generation is consumed and part is transferred to future generations to allow their consumption to rise.

2.6.2.2 Endogenous Human Capital

Adjusting the *unconstrained* solution above to allow for endogenous human capital is straightforward. Begin by re-introducing (2.25a), the efficiency condition that requires equal rates of return on human and physical assets. Let's assume a specific form for the human capital production function, $h_{t+1} = \Theta x_t^{\theta}$, where $\Theta > 0$, $0 < \theta < 1$. Now (2.25a) gives the following explicit solution for x_t

$$x_t = \left[\frac{\Theta\theta w}{R}\right]^{\frac{1}{1-\theta}}. \tag{2.34}$$

Notice again that, in the unconstrained case, the solution for human capital investment is simple. It involves equating the returns on the two assets and does not require any knowledge about the generational or dynastic wealth that determines consumption. Higher market wages raise the return to human capital investment, so investment must rise until the rate of return is driven back down to the interest rate. A rise in the interest rate causes a shift away from human capital investment, until its rate of return is driven up to the interest rate.

We complete the unconstrained solution by solving for consumption. Take the solution from (2.34) and define the new variable

$$\widehat{w} \equiv w\Theta\left[\frac{\Theta\theta w}{R}\right]^{\frac{\theta}{1-\theta}} - n\left[\frac{\Theta\theta w}{R}\right]^{\frac{1}{1-\theta}}, \tag{2.35}$$

which is the productively efficient lifetime wage income of the current generation minus the investment needed to ensure that the next generation also receives the efficient level of lifetime wages. Next, just substitute $\widehat{w}$ for w in the expression for W_∞ in (2.32). This substitution replaces the expression for wages assuming exogenous human capital with an expression for wages assuming endogenous human capital, net of the expenses of producing human capital across generations. After this substitution, the form of the solution for first period consumption of the current generation is the same as in the exogenous human capital case. Consumption of subsequent generations then follows from the generational Euler conditions given by (2.25b). Optimal financial transfer behavior will again be given by (2.33).

The *constrained* case is more difficult due to the interaction between household wealth and human capital investment. However, if we specialize a bit more, by assuming $\sigma = 1$, then an explicit solution is still possible. To obtain a solution, we use what is known as a *guess-and-verify* method. First, an educated guess about the form of the solution is made and then the solution is verified to satisfy the Bellman equation. It is similar to a situation where you think you know the answer to an algebraic equation and then you substitute the answer back into the equation to make sure the equation is satisfied.

In order to see the full logic, and the recursive approach used to find a solution, we relax the steady state assumption and allow r_t and w_t to vary across generations. Experience with models assuming logarithmic preferences suggests that a good guess for the value function is

$$V_t(W_t) = E + \left[\frac{1+\beta}{1-\beta\theta}\right]\ln W_t + \beta\left(\Phi_t^R + \left[\frac{1+\beta}{1-\beta\theta}\right]\Phi_t^w\right), \tag{2.36}$$

where E is an undetermined constant, $\Phi_t^R \equiv \sum_{i=1}^{\infty}\beta^{i-1}\ln R_{t+i-1}$ and $\Phi_t^w \equiv \sum_{i=1}^{\infty}\beta^{i-1}\ln w_{t+i}$. Notice that (2.36) bears a resemblance to the simple indirect utility function, associated with the pure life-cycle model without altruism, given in footnote 6. It is the experience in solving similar problems that forms the basis for making guesses in more complicated settings.

To verify that this does indeed represent a solution, go through the steps given in *Problem* 22. In working out the problem, you will find that the optimal solution for human capital investment is

$$x_t = \frac{\beta\theta W_t}{n}. \tag{2.37}$$

The constrained solution provides a sharp contrast to the unconstrained solution given by (2.34) for two reasons. First, human capital investment in children now depends on wealth. Second, the relevant wealth concept is once again the wealth of the *current* generation of parents and *not* the wealth of the entire family dynasty. Thus, despite the presence of altruistically motivated investments in children, the family is not connected as completely as in the unconstrained case. The wealth of the current household constrains investment in its children.

2.6.3 Warm Glow

Another approach to modeling intergenerational transfers is to assume that parents get utility, or a "warm glow," directly from *making* the transfers as opposed to getting utility from the *effects* of those transfers on their children's adult welfare. The warm-glow approach assumes that parents get *direct* satisfaction from taking care of their children and satisfying their sense of duty as good parents, or perhaps from the recognition they get from other adults in doing so. It also allows for the possibility that they enjoy giving one kind of transfer more than another. For example, they may prefer investing in their children's education more than giving them a saving bond, even when the effect on the child's wealth is the same in both cases.

An example of warm-glow preferences is given by the following utility function,

$$U_t = \ln c_{1t} + \beta \ln c_{2t+1} + \psi \ln h_{t+1} + \xi \ln b_{t+1} \tag{2.38}$$

where ψ and ξ are nonnegative preference parameters. The household values the human capital of children and the financial transfers to children, directly and independently.[10]

Under the warm-glow approach, the household maximizes (2.38) subject to the lifetime budget constraint associated with (2.18) and (2.19), and the human capital production function $h_{t+1} = \Theta x_t^{\theta}$, to generate the following equations for optimal transfers.

$$b_{t+1} = \frac{\xi}{1 + \beta + \theta\psi + \xi} \frac{W_t}{n} \tag{2.39a}$$

$$x_t = \frac{\theta\psi}{1 + \beta + \theta\psi + \xi} \frac{W_t}{n}. \tag{2.39b}$$

Both transfers are simply fractions of household wealth.

[10]Often, when the focus of the analysis is on intergenerational transfers, the life-cycle feature of the model is dropped. Households are modeled as living only a single period of adulthood in which they consume and make transfers to their children. We will see examples of these models in future chapters.

Contrasting (2.39a, 2.39b) to the transfer behavior when assuming altruism reveals several important differences. Households *always* make financial transfers—parents never find it optimal to give no financial transfers to their adult children. In addition, financial transfers are no longer affected by the household wealth of future generations in the family, as they were in (2.33). Under the warm-glow assumption, parents will not attempt to compensate for exogenous changes in their children's wealth.

The presence of financial transfers across generations does not imply that human capital investments are productively efficient as in (2.34). Depending on the wealth of the parents, human capital investments can be less than or *greater than* the efficient level. Some empirical researchers have concluded that marginal educational investments in most primary and secondary school students in the U.S. yield rates of return, in terms of future adult earnings, that are far below the rate of return on financial assets. This observation is inconsistent with the altruistic model, but not with the warm-glow model. As indicated, parents may value education for reasons beyond its impact on the wages of its adult children. If household wealth and the preference parameter ψ are sufficiently large, then the optimal human capital investment will exceed the productively efficient level, and the rate of return to educational investment will be below the rate of return on physical assets.

Finally, note that (2.37) and (2.39b) are quite similar. This is an important point. From the perspective of modeling human capital investments in children, assuming warm-glow preferences is similar to assuming altruism when the bequest constraint is binding. A binding bequest constraint is relevant because of the absence of intergenerational loan markets. Thus, warm-glow preferences are a convenient way of capturing the absence of intergenerational loans for human capital investment in children.

2.7 Related Literature

The first two-period overlapping generations model with physical capital and production was due to Diamond (1965). It has become one of the two workhorse models of macroeconomics. Unfortunately, most students are not exposed to this model, or any other model built on microeconomic foundations, as undergraduates. If the model does not sound familiar, you may want to read a more elementary undergraduate text along side of this book. An excellent intermediate undergraduate textbook treatment of the overlapping-generations model is provided by Auerbach and Kotlikoff (1998). Farmer and Schelnast (2013) provide a nice introductory graduate treatment of the overlapping-generations model, with a special focus on international trade. More advanced and more general treatments of the overlapping-generations model can be found in Azariadis (1993) and de la Croix and Michel (2002). These are important books for graduate students who want to concentrate on theoretical work.

The overlapping generation model was extended to include altruistic intergenerational transfers by Barro (1974), Drazen (1978) and Becker (1981, 1988). If the nonnegativity constraint on financial transfers is ignored, altruism provides a justification for infinitely lived agent model, the other workhorse model of macroeconomics. Barro (1997) provides an intermediate undergraduate treatment of macroeconomics using the infinitely-lived agent model. A more advanced undergraduate textbook that covers the overlapping-generations model and the extension to include intergenerational transfers is Lord (2001). Graduate treatments of the infinitely lived agent model include Romer (2001) and Acemoglu (2009).

A large empirical literature has developed to test the implications of altruistically motivated transfers. Most econometric tests indicate that either financial transfers are not motivated by altruism alone or that altruistic financial transfers are not operative for significant fractions of the population. See, for example, the papers by Wilhelm (1996) and Altonji et al. (1997). The finding that the intergenerational credit market is imperfect due to binding non-negativity constraints on financial transfers is consistent with empirical results suggesting that parent's lifetime resources affect the level of investments in their young children (Cunha and Heckman (2007)).

There are other reasons, not examined here, for intergenerational transfers. Some transfers are unintended or accidental. Individuals save for their own retirement, but die earlier than expected. The remaining wealth is often left to children. For analysis and applications of models with unintended transfers, start with articles by Davies (1981), Abel (1985), and Hurd (1989). In addition, some transfers may be intentionally made for strategic reasons. In particular, parents might accumulate wealth that is potentially bequeathed to children in order to elicit services from them when they are adults. For more reading on strategic bequests, begin with Bernheim et al. (1985) and Cox (1987).

Calibrated dynamic general equilibrium models were first used in public finance to examine the effects of tax reform (Summers (1981) and Auerbach et al. (1983)) and in macroeconomics to explain business cycles (Kydland and Prescott (1982)). Calibration methods have since been extended to every area of macroeconomics. For a general discussion of the approach, including additional applications, see Kydland and Prescott (1996), followed by a critique from Hansen and Heckman (1996).

The calibration experiment we presented in Sect. 2.4, that uses the model with physical capital to explain economic growth, is based on King and Rebelo (1993). The extension to include human capital in Sect. 2.5 is based on Rangazas (2000, 2002). Cordoba and Ripoll (2013) and Manuelli and Seshadri (2014) offer quantitative theories where human capital accumulation is the dominant determinant of an economy's labor productivity. An alternative approach that stresses the connection between TFP and *broad notions* of capital, emphasizing the knowledge embodied in firms rather than individuals, has been developed by Parente and Prescott (2000).

2.8 Exercises

Questions

1. Define the following concepts and give an example of each.
   (a) *technology*
   (b) *capital*
   (c) *physical capital*
   (d) *human capital*
2. Explain the meaning of (2.2a) and (2.2b). What variables are determined by these two equations?
3. Let's relate the discussion of the firm to something you know well. Think of the physical capital stock as fixed, as in the short-run model of the competitive firm from introductory and intermediate microeconomics. Sketch the marginal product of labor as a function of the employment level of a firm. Next, add the competitive market wage rate to the diagram. Finally, locate the firm's profit-maximizing employment level. How does the profit-maximizing demand for labor change if there an increase in the firm's capital stock? An increase in A? An increase in the market wage? In forming your answer, only think about how an *individual* firm's demand for labor would change.
4. Explain the differences between *rental rate*, *rate of return on capital*, and *interest rate*.
5. Explain how each of the following affects current consumption, future consumption, and saving: (a) wages, (b) return to capital, and (c) the preference parameter β.
6. How is Fig. 2.1 altered by an increase in w_{t-1}? An increase in n? Explain your answers intuitively.
7. Explain why an increase in this period's capital stock causes an increase in next period's capital stock. Why does the linkage become weaker as the economy accumulates more capital?
8. As the economy moves toward the steady state from below what happens to its growth? What happens to the growth rate if it heads to the steady state from above?
9. Suppose that economy A and economy B have identical structures (production technologies, preferences, and population growth rates), but economy A has a higher capital-labor ratio. Which country is "richer"? Which country grows faster? Your answer explains what is known as *conditional convergence*. Show, by example, that if two countries do *not* have identical structures that *absolute* or *unconditional convergence* is not guaranteed. See Sect. 3.8 of Chap. 3 for more discussion of convergence concepts.
10. Why is it clear that the model must include technological progress in order to match empirical data?
11. Explain why a high rate of population growth will make a country poor.

12. How does an increase in exogenous technological change affect capital intensity, interest rates, and labor productivity? In what way is it similar to the effects of an increase in population growth and in what way is it different?
13. Let the annualized value of d be defined as d^a. Assume that $d^a = 0.01$.
    (a) What is d if each period in the model last 30 years? (Hint: recall how r_t and r_t^a are related from *Problem* 3)
    (b) What is the annualized growth rate of worker productivity, Y_t/M_t, in the steady state? In thinking about this question it may help to refer to the results of *Problem* 16.
    (c) What can you say about the annualized growth rate in worker productivity as the economy approaches the steady state from below?
14. Explain what it means to calibrate a model. Briefly describe the calibration of the model of physical capital accumulation in Sect. 2.4. Mention how each parameter was set. How was the initial value of k determined?
15. Discuss the design and the results of the calibration experiment when the model of physical capital accumulation was used to historically replicate growth in the U.S. from 1870 to 1990.
16. What is human capital? What are the two inputs used to produce human capital in Sect. 2.5? If $\theta_1 = 0.10$ and $\theta_2 = 0.40$, what is the ratio of human capital per worker in 1990 relative to 1870?
17. Discuss the design and the results of the calibration experiment where the model of physical capital accumulation was extended to include human capital.
18. Give an example of a human capital transfer and a financial transfer. When would a parent's payment of college tuition be equivalent to a financial transfer to the child? Do government-guaranteed college loans alleviate the effects of the legal restriction that parents cannot leave debt to their children?
19. If intergenerational transfers are altruistically motivated, explain why human capital investments are productively efficient if financial transfers are positive, but are inefficiently low if financial transfers are zero.
20. Suppose that intergenerational transfers are unconstrained. Intuitively guess what happens to c_{1t} and c_{1t+1} when R_t increases. To make your guess note that (2.25b) has exactly the same form as the expression for the optimal ratio of future to current consumption over the life-cycle (see 2.22a, 2.22b and *Problem* 6). Suppose now that households are constrained. Based on analogous reasoning used to make guesses about the effects of an increase in R_t, use (2.26b) to guess what happens to c_{1t}, c_{1t+1}, and x_{t+1} when w_{t+1} increases.
21. Under special assumptions, Eq. (2.37) gives an explicit solution for constrained human capital investments. Use your answer to *Question* 20 to explain why w_{t+1} does not enter the equation.
22. Assume that intergenerational transfers are altruistically motivated and that financial transfers are positive. Explain what happens to human capital investments in children and financial transfers if.
    (a) W_t increases by one unit and W_∞ is held constant.
    (b) W_∞ increases by one unit and W_t is held constant.

    Repeat the exercise when (i) financial transfers are zero and then again when (ii) transfers are motivated by warm glow.

23. Explain why (2.37) and (2.39b) are similar despite the fact that they are generated by two different assumptions about intergenerational transfers.
24. Under each of the motives for intergenerational transfers explain if and when investments in human capital can be greater than the productively efficient level.
25. Based on the results of this chapter, offer an explanation for the stylized growth facts, *G1* and *G5*.

Problems

1. Show that (2.1) exhibits the neoclassical properties of diminishing marginal productivity and constant returns to scale.
2. Derive Eqs. (2.2a, 2.2b) and (2.3a, 2.3b). How are they related?
3. The time periods of the model are rather abstract, representing an entire working life, say 30 years. We interpret r_t as the compounded rental rate earned over the 30 year period. We can think of an *annualized* rental rate, r_t^a, by introducing the definition, $1 + r_t = \left(1 + r_t^a\right)^{30}$.

   Suppose that the annualized return to capital is 7.4% or $r_t^a = 0.074$. Assuming that $A = 1$, and $\alpha = 1/3$, find the numerical values for the following variables that are consistent with a perfectly competitive equilibrium given this particular value for r_t^a
   (a) r_t
   (b) k_t
   (c) w_t
   (d) economic profit

   Assuming the same value for r_t^a, redo the calculations if $A = 30$. Intuitively explain the effect on (b)-(d) of assuming the higher value for A.
4. What is total income in the model with capital and production? Show that the value of output is equal to the value of income.
5. Derive the optimal life-cycle behavior given by (2.5a, 2.5b) and (2.6).
6. Sketch the lifetime budget constraint of a household, $c_{1t} + \frac{c_{2t+1}}{R_t} = w_t$, with the two choice variables, c_{1t} plotted on the horizontal axis and c_{2t+1} on the vertical axis. What happens to the diagram if R_t increases? Conceptually decompose how the diagram is affected by an increase in R_t in terms of (a) opportunities for consumption in both periods and (b) the opportunity cost of current consumption in terms of forgone future consumption. Intuitively think about how (a) and (b) affect the optimal choice of consumption in the first period—the effect of (a) is called the *income* or *wealth effect* and the effect from (b) is called the *substitution effect*. Why are the names for the effects appropriate? What must be true about these two conceptual effects of an increase in R_t to be consistent with the optimal choice of c_{1t} given in (2.5a)?
7. We can generalize household preferences by using a *Constant Elasticity of Substitution* (CES) utility function,

$$u_t = U(c_{1t}, c_{2t+1}) = \frac{\left(c_{1t}^{1-1/\sigma} - 1\right) + \beta\left(c_{2t+1}^{1-1/\sigma} - 1\right)}{(1 - 1/\sigma)}.$$

The new parameter is the *intertemporal elasticity of substitution* (σ). The intertemporal elasticity of substitution is a measure of the individual's willingness to substitute current for future consumption when the relative price of future consumption falls. Subtracting 1 from each consumption term is done for purely technical reasons. It allows the commonly used logarithmic utility function, $\ln c_{1t} + \beta \ln c_{2t+1}$, to appear as a special case when $\sigma = 1$ (see the Technical Appendix section A.5).

Show if households maximize the CES lifetime utility function subject to the lifetime budget constraint, the solution gives us the following optimal consumption and saving behavior.

(a) $c_{1t} = \Psi_{1t} w_t$
(b) $c_{2t+1} = \Psi_{2t} w_t$
(c) $s_t = \Psi_{1t}\beta^{\sigma} R_t^{\sigma-1} w_t$,

$$\text{where } \Psi_{1t} \equiv \frac{1}{1 + \beta^{\sigma} R_t^{\sigma-1}} < 1 \text{ and } \Psi_{2t} \equiv \frac{\beta^{\sigma} R_t^{\sigma}}{1 + \beta^{\sigma} R_t^{\sigma-1}}.$$

The relative strength of the income and substitution effects identified in *Problem* 6 is determined by σ. The greater is the value of σ the stronger is the substitution effect and the weaker is the income effect. When $\sigma = 1$, the income and substitution effects cancel exactly and the saving rate become a constant fraction of wages.

8. Starting from the definition of k^s, derive Eq. (2.7).
9. Show that if we use the saving behavior with the CES preferences from *Problem* 7 that the transition equation becomes

$$k_t = \frac{(1-\alpha)Ak_{t-1}^{\alpha}}{n}\left[\frac{1}{1 + \beta^{-\sigma}(1 + r_t - \delta)^{1-\sigma}}\right].$$

10. Starting with the adjusted definition of k^s, derive the transition Eq. (2.9).
11. How do we know that the transition equation will be concave in Fig. 2.2? The concavity of the transition function establishes three crucial properties of the steady state equilibrium: (i) existence (there is a steady state), (ii) uniqueness (there is only one steady state with $k > 0$), and (iii) dynamic stability (if you start away from the steady state you will always move toward it). Use the diagram to explain this.
12. *Transition Paths I*

Under the following parameter assumptions: $A = n = 1$, $d = 0$, $\beta = 1/2$, $\alpha = 1/3$, and an initial capital intensity of $k_0 = 0.0500$, compute the values of k_t over the next 5 periods using (2.9). What is the exact value of k_t in the steady state?

13. *Transition Paths II*
 Use the same assumptions as in *Problem* 12, but now let $A = 10$. Compute the values of k_t over the next 5 periods and in the steady state. Use a transition equation diagram to contrast the solutions to *Problems* 12 and 13.
14. *Transition Paths III*
 Use the same assumptions as in *Problem* 13, but now let $n = 1.5$. Compute the values of k_t over the next 5 periods and in the steady state. Use a transition equation diagram to contrast the solutions to *Problems* 13 and 14.
15. *Transition Paths IV*
 Use the same assumptions as in *Problem* 13, but now let $d = 0.5$. Compute the values of k_t over the next 5 periods and in the steady state. Use a transition equation diagram to contrast the solutions to *Problems* 13 and 15. What is the difference between Problems 14 and 15?
16. In the model, adjusted for technological progress, the production function becomes $Y_t = AK_t^{\alpha}H_t^{1-\alpha}$.
    (a) If we redefine k_t as $k_t \equiv K_t/H_t$, show $Y_t = Ak_t^{\alpha}H_t$.
    (b) Show worker productivity is $Y_t/M_t = Ak_t^{\alpha}D_t$
17. Using a calculator or a computer, reproduce the values associated with the historical simulation displayed in Fig. 2.3. Note that the period growth factor, one plus the period growth rate, is $\frac{(Y/M)_{t+1}}{(Y/M)_t} = \left(\frac{k_{t+1}}{k_t}\right)^{\alpha}\frac{D_{t+1}}{D_t}$. To get the annual growth rates in the figure you must calculate *annualized* value of the period growth factor.
18. Use the results of *Problems* 7 and 9 to derive the transition for k_t when $\sigma \neq 1$ and there is exogenous changes in H_t through changes in D_t. Redo the numerical exercise in Problem 17 when $\sigma = 0.50$. Note, using a similar approach to that described in the text, you will have to select a new initial value of k_1 to generate the required total growth by period 5. Explain the difference in the paths of k_t when $\sigma = 0.50$ and when $\sigma = 1$.
19. Use the results of *Problem* 7 and let $\delta = 1$ and $\sigma = \frac{2-\alpha}{1-\alpha}$. Derive an explicit transition equation for k_t in terms of k_{t-1}. Sketch the transition equation as accurately as you can.
20. Derive the financial transfer equations under the two motives for intergenerational transfers. In particular, use (2.22b), (2.27a), and (2.32) to derive (2.33). Next, maximize (2.38) subject to the household lifetime budget constraint to get (2.39a).
21. Assuming that (i) intergenerational transfers are altruistically motivated and (ii) the non-negativity constraint on financial transfers does not bind, derive *aggregate* consumption in period-t in terms of the behavior of a "representative agent," whose consumption behavior is a function of the dynasty's wealth. First note that aggregate consumption in period-t is $C_t \equiv N_t c_{1t} + N_{t-1}c_{2t} = N_t\left[c_{1t} + \frac{c_{2t}}{n}\right]$. Next use (2.22b) and (2.25b) to get $C_t = \kappa N_t c_{1t}$, where $\kappa \equiv 1 + n^{\sigma-1}$ and c_{1t} is given by (2.32). This shows that aggregate consumption is proportional to the consumption of a single representative agent, the current young adult of the family dynasty.

22. Take a deep breath and complete the following steps to verify that (2.36) and (2.37) are indeed the solution to the constrained-problem with human capital when $\sigma = 1$.
    (i) Write (2.36) for generation-$t + 1$ instead of generation-t.
    (ii) Substitute your answer to (i) for $V(W_{t+1})$ in (2.21).
    (iii) Use the relationship between future and current life-cycle saving (see, for example, 2.5a, 2.5b) to write U_t solely in terms of c_{1t} and then in terms of W_t and x_t by using the lifetime budget constraint associated with (2.18) and (2.19) (remember that $\sigma = 1$ and $b_{t+1} = 0$)
    (iv) Now you have the objective function in (2.21) written solely in terms of the choice variable x_t. Show the first order condition for maximizing (2.21) with respect to x_t yields (2.37).
    (v) Finally, substitute (2.37) back into the objective function from (iv) and show that $V(W_t)$ is in fact (2.36). This is done by organizing all the constant terms into a term labeled E, and then forming three other expressions that group terms involving W_t, human capital rental rates, and interest rates. Careful, this is probably the most difficult part.
23. An alternative way of thinking about equilibrium is to focus on the goods market rather than the capital market. In the simple model of physical capital accumulation, with no human capital or technological progress, the goods market clearing condition just says that the supply of goods or output, Y_t, must equal the demand for consumption and investment goods, $C_t + K_{t+1} - (1 - \delta)K_t$, so that

$$C_t + K_{t+1} - (1 - \delta)K_t = Y_t,$$

where, as in *Problem* 21, you must think of C_t as the total consumption of all generations. We can write this as a transition equation in the capital-labor ratio by dividing by M_t and rearranging a bit to get

$$nk_{t+1} = Ak_t^{\alpha} - (1 - \delta)k_t - \frac{C_t}{M_t}.$$

Show that in the two period overlapping generation model, this condition is equivalent to the transition Eq. (2.8) that was derived from the capital-market equilibrium condition.

Appendix

A—Derivative of the Value Function

Substituting (2.22a, 2.22b) back into U_t gives

$$U_t = \frac{[\Psi_{1t}(W_t - n(x_t + b_{t+1}))]^{1-1/\sigma}}{1 - 1/\sigma} + \beta \frac{[(\beta R_t)^{\sigma} \Psi_{1t}(W_t - n(x_t + b_{t+1}))]^{1-1/\sigma}}{1 - 1/\sigma}$$
$$= \Psi_{1t}^{-1/\sigma} \frac{(W_t - n(x_t + b_{t+1}))^{1-1/\sigma}}{1 - 1/\sigma}.$$

Next, denote the optimal choices of the intergenerational transfers with an asterisk so that we can write

$$V_t(W_t) = \Psi_{1t}^{-1/\sigma} \frac{\left(W_t - n\left(x_t^* + b_{t+1}^*\right)\right)^{1-1/\sigma}}{1 - 1/\sigma} + \beta V_{t+1}\left(W_{t+1} H\left(x_t^*\right) + (1 + r_{t+1}) b_{t+1}^*\right).$$

Note that the optimal intergenerational transfers are in general functions of W_t (in the constrained-case, financial transfers are zero and thus do *not* depend on W_t). However, because the derivative of the right-hand-side with respect to these choices is zero, then we may ignore these indirect effects of W_t when differentiating V_t with respect to W_t (in the constrained-case, the derivative is taken only with respect to human capital investments). Thus, the derivative of the value function with respect to W_t*includes only the direct effect*, working through U_t in the first expression above. This result is a general one, and is known as the *envelope theorem*. Equation (2.24) is the envelope theorem applied to $V_{t+1}(W_{t+1})$.

B—Many-Period Models

In models that extend beyond two periods, it is easier to proceed by thinking of the equilibrium or market clearing condition in terms of goods rather than capital. The two ways of thinking about things are actually equivalent, but the exposition is easier if we conduct the discussion in terms of goods (see *Problem* 23). This is especially true if one wants to contrast the infinitely-lived agent approach with the overlapping generations' approach. To reduce notation, we will limit the discussion to the simple model of physical capital accumulation with no human capital or technical progress. In addition, the key points can be made for the case with $\sigma = 1$.

As shown in *Problem* 23, using the goods market approach allows the transition equation for the economy to be written as

$$nk_{t+1} = Ak_t^{\alpha} - (1 - \delta)k_t - \frac{C_t}{M_t}.$$

In the overlapping-generations model with two period lifetimes, this equation reduces nicely to a first-order difference equation (*Problem* 23). This is because the right-hand-side can be reduced to the consumption behavior of a single generation whose consumption depends only on period-t wages and thus only on k_t. The old generation consumes all its income. As a result, their income and their consumption are equal and cancel from the right-hand-side of the transition equation.

However, with more periods of life there will be more generations on the right-hand-side, whose income is less than their consumption, i.e. who save by purchasing capital. The consumption behavior of these generations will depend on wages earned *before* period-t. For example, think of a model with five periods of life—four working periods and one retirement period. The aggregate consumption behavior on the right-side sums over five different generations. In general, it is only the generation who is retired that will "cancel out." The saving of other generations will generally not be zero, so all variables that affect their consumption behavior will appear in the transition equation.

Consider the next oldest generation, who is in the last period of its working life. This generation's working life began three periods ago. Their consumption behavior in period-t will depend on the wages they earned in each of those periods. Thus, wages as far back at period $t-3$ will appear in the transition equation. The wages in each period are determined by the capital-labor ratio from that period. This implies that the transition equation will be a *fifth-order* difference equation, including the variables, k_{t+1}, k_t, k_{t-1}, k_{t-2}, k_{t-3}. The important point is that the *state variables*, here the capital-labor ratios, characterizing the economy increase as the number of periods of the life-cycle increase. This *curse of dimensionality* raises the computational complexity of the model when the number of periods in the life-cycle is large.

In contrast, if one assumes that the generations are linked by intergenerational financial transfers, the transition equation of the economy is a *second order* difference equation, *no matter how many periods in the life-cycle* are included. To see this, first note that as long as financial intergenerational transfers link the generations together, then we can write $C_t = \kappa N_t c_{1t}$ (see *Problem* 21). As periods in the life-cycle are added, only the value of κ changes. For example, $\kappa = 2$, with two periods of life, and $\kappa = 5$, if there are five periods of life.

Next, solve for c_{1t} using the goods-market version of the transition equation to get

$$c_{1t} = \frac{M_t}{N_t \kappa}\left[Ak_t^{\alpha} - (1-\delta)k_t - nk_{t+1}\right].$$

Now substitute into (2.25b) to get

$$\left[Ak_{t+1}^{\alpha} - (1-\delta)k_{t+1} - nk_{t+2}\right] = \left[\frac{\beta\left(1+\alpha Ak_{t+1}^{\alpha-1} - \delta\right)}{n}\right]\left[Ak_t^{\alpha} - (1-\delta)k_t - nk_{t+1}\right],$$

This is a *second-order* difference equation in k_t that is completely *independent of the number of periods in the life-cycle*. Thus, the infinitely-lived agent simplification is able to avoid the dreaded curse of dimensionality. A very nifty and useful result.

The lesson is that if you want to do computations, use the infinitely-lived agent approach whenever you can get away with it (e.g. business cycle analysis). Unfortunately, the evidence suggests that for the issues we examine in this book, those that directly focus on private intergenerational transfers and government policies that transfer resources across generations, one must stick with the overlapping generations approach.

References

Abel A (1985) Precautionary saving and accidental bequests. American economic review 75 (4):777–791.
Acemoglu D (2009) Introduction to modern economic growth. Princeton University Press, Princeton, New Jersey.
Altonji J, Hayshi F, Kotlikoff L (1997) Parental altruism and inter vivos transfers: Theory and evidence. Journal of political economy 105(6):1121–1166.
Auerbach A, Kotlikoff L (1998) Macroeconomics: An integrated approach. MIT Press, Cambridge, Massachusetts.
Auerbach A, Kotlikoff L, Skinner J (1983) The efficiency gains from dynamic tax reform. International economic review 24(1):81–100.
Azariadis C (1993) Intertemporal macroeconomics. Blackwell, Oxford, UK.
Barro R (1974) Are government bonds net wealth? Journal of political economy 82(6):1095–1118.
Barro R (1997) Macroeconomics. MIT Press, Cambridge, Massachusetts.
Becker G (1974) A theory of social interactions. Journal of political economy 82(6):1063–1094.
Becker G (1981) A treatise on the family. Harvard University Press, Cambridge, Massachusetts.
Becker G (1988) Family economics and macro behavior. American economic review 78(1):1–13.
Bernheim D, Shleifer A, Summers L (1985) The strategic bequest motive. Journal of political economy 93(6):1045–1076.
Córdoba J, Ripoll M (2013) What explains schooling differences across countries? Journal of monetary economics 60(2):184–202.
Cox D (1987) Motives for private income transfers. Journal of political economy 95(3):508–546.
Davies J (1981) Uncertain lifetime, consumption and dissaving in retirement. Journal of political economy 89(3):561–577.
de la Croix D, MP (2002) A theory of economic growth. Cambridge University Press, Cambridge.
Diamond P (1965) National debt in a neoclassical growth model. American economic review 55 (5):1126–1150.
Drazen A (1978) Government debt, human capital, and bequests in a life-cycle model. Journal of political economy 86(3):505–516.
Farmer K, Schelnast M (2013) Growth and international trade: An introduction to the overlapping generations approach. Springer-Verlag, Berlin.
Gollin D (2002) Getting income shares right. Journal of political economy 110(2): 458–474.
Hansen L, Heckman J (1996) The empirical foundations of calibration. Journal of economic perspectives 10(1):87–104.
Heckman J, Cunha F (2007) The technology of skill formation. American economic review 97(2):31–47.
Hurd M (1989) Mortality risk and bequests. Econometrica 57(4):779–813.
King R, Rebelo S (1993) Transitional dynamics and economic growth in the neoclassical growth model. American economic review 83:908–931.
Kocherlakota N (1996) The equity premium: It's still a puzzle. Journal of economic literature 34 (1):42–71.
Kydland F, Prescott E (1982) Time to build and aggregate fluctuations. Econometrica 50(6): 1345–1370.
Kydland F, Prescott E (1996) The computational experiment: An econometric tool. Journal of economic perspectives 10(1):69–86.
Lord W (2001) Household dynamics, policies, and economic growth. Oxford University Press, New York.
Lord W, Rangazas P (2006) Fertility and development: The roles of schooling and family production. Journal of economic growth 11(3):229–261.
Manuelli R, Seshadri A (2014) Human capital and the wealth of nations. American economic review 104 (9):2736–2762.

Mourmouras A, Rangazas P (2009) Reconciling Kuznets and Habbakuk in a unified growth model. Journal of economic growth 14(2):149–181.

Parente S, Prescott E (2000) Barriers to riches. MIT Press, Cambridge, Massachusetts.

Rangazas P (2000) Schooling and economic growth: A King-Rebelo experiment with human capital. Journal of monetary economics 46(2):397–416.

Rangazas P (2002) The quantity and quality of schooling and U.S. labor productivity growth (1870–2000). Review of economic dynamics 5(4): 932–964.

Romer D (2001) Advanced macroeconomics. McGraw-Hill, New York.

Stokey N, Lucas R (1989) Recursive methods in economic dynamics. Harvard University Press, Cambridge Mass.

Stokey N, Rebelo S (1995) Growth effects of flat-rate taxes. Journal of political economy 103 (3):519–50.

Summers L (1981) Capital taxation and capital accumulation in a life-cycle growth model. American economic review 71(4):533–544.

Wallis J (2000) American government finance in the long-run: 1790 to 1990. Journal of economic perspectives 14(1):61–82.

Wilhelm M (1996) Bequest behavior and the effect of heirs' earnings: Testing the altruistic model of bequests. American economic review 86(4):874–92.

Fiscal Policy

3

This chapter highlights the important role played by the government in jump starting and sustaining economic growth. The government provides the country's infrastructure or *public capital*—laws, roads, education, public health, and utilities—that secure property rights and raise the productivity of private capital. Growth in public capital keeps the marginal return to private capital from falling dramatically during industrialization, helping to explain *G1*. The need for government to lay the foundation for growth helps explain the vast differences in experiences of developing countries after WWII, summarized in *G6*. Growth Miracles and Growth Disasters are largely driven by especially good and especially bad government policy.

3.1 Introducing the Government

We now introduce fiscal policy into the model to study its impact on economic growth. For simplicity, assume a single source of government revenue; a tax rate levied on wage income (τ_t). On the spending side, we focus on *government purchases*. Government purchases include the purchase of privately produced goods, both consumption and investment goods, and the employment of government workers that provide public services.

The most common situation analyzed in macroeconomics is where the government purchases a private consumption good that is used by government employees (C_t^g). For concreteness, one can think of direct consumption purchases as food, uniforms, weapons, and equipment provided to the military.

Of particular importance to our analysis is government capital such as roads, ports, public utilities, and public schools, or even the intangible capital comprised of the nation's laws that protect property rights (G_t). Recall from Sect. 2.4 of Chap. 2, that even when we attempted to calibrate the *annual* depreciation rate on private capital realistically, the *period* depreciation rate in our model was close to one. We set $\delta = 1$ throughout and also assume that government capital depreciates completely

S. Das et al., *Economic Growth and Development*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-319-89755-4_3

after one period. This means that choosing next period's public capital stock is the same as choosing this period's government investment in public capital.

Finally, there are the wages paid to government officials (w_tD_t). The government work force is a fraction (ς) of the private work force of young households that are employed by firms. The total number of young households in period t is now $(1 + \varsigma)$ N_t. Other than the fact that the government officials work for the government rather than for private firms, they are identical to private households; earning the same wage and possessing the same preferences. Wages paid to government employees are recorded as a ond type of government consumption purchase.

The sources and uses of funds are summarized and connected via the *government budget constraint* for period *t*

$$\tau_t w_t D_t (1 + \varsigma) N_t = C_t^g + w_t D_t \varsigma N_t + G_{t+1}. \tag{3.1}$$

The left-hand-side is the total revenue from the wage tax. The right-hand-side presents the uses of funds: the two types of government consumption and investment in public capital.

3.1.1 Government Capital and Private Production

We need to model the contribution of government capital to private production. As before, the Cobb-Douglas production function is

$$Y_t = AK_t^{\alpha} (H_t)^{1-\alpha}, \tag{3.2}$$

where $H_t = D_tM_t = D_tN$. For notational simplicity, no population growth is assumed in the next two sections ($n = 1$). However, now the productivity index (D) is a function of exogenous disembodied technology (E) *and* endogenous public capital per worker ($G/((1 + \varsigma)N)$) and is given by

$$D_t = E_t^{1-\mu} (G_t / ((1 + \varsigma)N))^{\mu}, \tag{3.3}$$

where $0 < \mu < 1$ is a constant parameter that determines the effect of public capital on output in the same manner as α does for private capital. We assume that E progresses at the exogenous rate q. In addition to this exogenous technological progress, public infrastructure raises the productivity of the private sector. Public education and public health raise worker productivity. Roads and property right protection free resources for production that otherwise would have spent commuting to work, delivering intermediate and final goods, and securing assets.

It is most descriptively accurate to assume that public workers also draw on the services of public infrastructure. For this reason, they also contribute to the dilution of public capital and lower the value of $G/E(1 + \varsigma)N$. To keep the notation simple we will sometimes assume that public capital is only used by private workers when it does not affect the key points we are trying to make.

Physical capital intensities are now defined per worker that uses the capital after adjusting for the exogenous source of productivity, E. We define public capital intensity as $g \equiv G/E(1+\varsigma)N$ and private capital intensity as $k \equiv K/EN$. The full productivity index given by (3.3) can then be written as $D_t = E_t g_t^{\mu}$.

With the new definitions of physical capital intensity, output per worker is

$$Y_t/(1+\varsigma)N = AK_t^{\alpha}\left(E_t g_t^{\mu} N\right)^{1-\alpha}/(1+\varsigma)N = AE_t g_t^{\mu(1-\alpha)} k_t^{\alpha}/(1+\varsigma).$$

There are now five determinants of worker productivity, three from before and two new ones:

- A unmeasured features of an economy that do not change on a regular basis but that affect the level of productivity (natural resources, climate, reliance on markets vs a command economy approach to allocate resources)
- E_t state of technology or knowledge about production that evolves steadily over time (firm organization, production methods, machine design)
- g_t publicly provided capital and infrastructure (roads, public education, property right protection)
- k_t private physical capital (plant and equipment)
- ς relative size of public sector employment that may provide unmeasured services but does not directly raise private production (soldiers, bureaucrats, public officials)

Firms continue to operate in perfectly competitive factor and output markets. As before, they rent physical capital and effective workers to produce output and maximize profits, $Y_t - w_t H_t - r_t K_t$. The profit-maximizing factor mix must satisfy

$$r_t = \alpha A\left(E_t g_t^{\mu} N\right)^{1-\alpha} K_t^{\alpha-1} = \alpha A g_t^{\mu(1-\alpha)} k_t^{\alpha-1} \tag{3.4a}$$

$$w_t = (1-\alpha)A\left(E_t g_t^{\mu} N\right)^{-\alpha} K_t^{\alpha} = (1-\alpha) g_t^{-\mu\alpha} A k_t^{\alpha}, \tag{3.4b}$$

Note that the complementary nature of public capital as a productive input means the marginal product of capital is increasing in g_t, roads increase the productivity of private inputs. If k_t and g_t both rise over the course of development, there need not be a marked decline in the return to private capital. Also, remember w_t is the rental rate paid to a unit of *effective* labor. The full wage paid to an *actual* worker is $w_t D_t = (1-\alpha)AE_t g_t^{\mu(1-\alpha)} k_t^{\alpha}$. In summary, on the production side of the economy, the key new feature is that public capital positively affects the marginal product of private inputs and factor prices.

3.1.2 Households and Taxes

There are $(1+\varsigma)N$ young households in each period. The households are standard two-period life-cycle savers with the same preferences as before. With fiscal policy they now face a wage tax rate (τ) when young. The household's lifetime budget constraint is given by

$$c_{1t} + \frac{c_{2t+1}}{R_t} = (1 - \tau_t) w_t D_t. \tag{3.5}$$

Maximizing the log utility function from Chap. 2, subject to (3.5), yields the following consumption and saving functions in the presence of taxes

$$c_{1t} = \frac{(1 - \tau_t) w_t D_t}{1 + \beta} \tag{3.6a}$$

$$c_{2t+1} = \beta R_t c_{1t}. \tag{3.6b}$$

$$s_t = \frac{\beta(1 - \tau_t) w_t D_t}{1 + \beta}. \tag{3.6c}$$

Taxes reduce lifetime wealth and consumption. Taxes also reduce first period income flows used to finance saving.

3.1.3 Capital Market Equilibrium and Fiscal Policy

Beyond the fact that the wage tax lowers household saving, the form of the capital market equilibrium condition remains unaffected,

$$K_{t+1} = (1 + \varsigma) N_t s_t. \tag{3.7}$$

3.2 Government Purchases

We are ready to think about how government purchases affect economic growth. The two types of expenditures are examined separately: first government consumption purchases are analyzed and then we move on to government investment purchases. When examining consumption purchases, we further simplify notation by assuming no technological progress: $E_t \equiv 1$ and $q = 0$.

3.2.1 Government Purchases–Consumption of Private Goods

First further simplify by assuming $\varsigma = \mu = 0$ to eliminate discussion of public employment and government investment. If all the government does is purchase private consumption goods, the government budget constraint is $\tau_t w_t N = C_t^g$.

The taxes paid by each worker can then be written as $\tau_t w_t = c_t^g$, where $c_t^g \equiv C_t^g/N$, government consumption per household.

The capital market equilibrium condition is

$$K_{t+1} = N\frac{\beta}{1+\beta}(1-\tau_t)w_t. \tag{3.8}$$

After substituting for $\tau_t w_t$ and dividing by N, the transition equation for private capital intensity is

$$k_{t+1} = \frac{\beta}{1+\beta}\left[(1-\alpha)Ak_t^\alpha - c_t^g\right]. \tag{3.9}$$

An increase in government consumption purchases raises the wage tax and lowers private savings. The decline in savings lowers the funding for purchases of private capital and the transition equation shifts down, resulting in less economic growth and lower steady state levels of k_t and y_t. Government consumption might be needed to provide important services—soldiers need to be fed and equipped to provide national defense—but there is a cost in lower worker productivity.

3.2.2 Government Purchases–Consumption Via Public Employment

The payment to government officials for their services is a second type of government consumption purchase. If we set $\varsigma > 0$and only examine government employment, the government budget constraint becomes, $\tau_t w_t(1 + \varsigma)N = w_t\varsigma N$, so $\tau_t = \tau = \varsigma/(1 + \varsigma)$. The tax rate simply reflects the relative size of the government employment share.

The capital market equilibrium condition is again given by (3.8). After substituting for τ and dividing by N, the transition equation with government employment purchases is

$$k_{t+1} = \frac{\beta}{1+\beta}(1-\alpha)Ak_t^\alpha. \tag{3.10}$$

Equation (3.10) is precisely the transition equation found in an economy with no government. The presence of a government sector that absorbs the economy's labor does not affect the capital-labor ratio or the productivity of workers in the private sector. An increase in ς does raise the wage tax, which lowers saving of private households and private capital accumulation. This effect is offset by the fact that the tax revenue is used to pay public sector workers who save at the same rate as private households. There is no change in national saving.

However, the government does divert labor from private sector production. Total output is $Y_t = y_t N = Ak_t^\alpha N$, so output per worker in the *economy as a whole* is $Y_t/(1+\varsigma)N = Ak_t^\alpha/(1+\varsigma)$. Output and income per capita falls as the government sector becomes relatively larger, as does the after-tax wealth of households, $(1 - \tau) w_t = w_t/(1 + \varsigma)$. These results depend on the government workers not being directly productive or at least their production is not measured as output. Think again of soldiers that do not produce goods directly but that provide unmeasured protection services for the country. The more soldiers used to provide protection services, the fewer private goods are available per young household.

Recall from introductory economics that national income accounting attempts a crude measure of the value of untraded government services by using the wages paid to public employees. Our concept of output only includes private goods because public officials in the model do not directly provide any services. Thus, in our model the higher is ς, the lower is private output per worker (private and public).

3.2.3 Government Consumption Purchases

If we combine our results so far we can draw conclusions about the combined cost of both types of government consumption, e.g. the cost of hiring, feeding, and equipping the army. We know the private consumption good purchases needed to feed and equip the army will raise the nation's consumption rate, lower savings and private capital accumulation, as well as future worker productivity. In addition, hiring soldiers diverts the work force away from the production of private goods. So, government consumption causes the country to have fewer, and less productive, workers devoted to making private goods. In providing national defense, as with all types of government consumption, one needs to account for the costs in terms of forgone private good production and consumption—the "guns and butter" trade-off.

3.2.4 Government Purchases–Investment Goods

Now let's ignore government consumption and focus on government investment. To re-introduce the effect of public capital on D_t, assume $\mu > 0$. With government investment purchases only, the government budget constraint becomes $\tau_t w_t D_t N = G_{t+1}$, which implies

$$g_{t+1} = \tau_t w_t D_t. \tag{3.11}$$

The capital market equilibrium condition yields the following transition equation

$$k_{t+1} = \frac{\beta}{1+\beta}(1 - \tau_t) w_t D_t, \tag{3.12}$$

similar to before, except now there is the presence of D_t, which is a function of g_t.

Combining (3.11) and (3.12), we see that public capital is proportional to private capital,

$$g_{t+1} = \frac{\tau_t}{1-\tau_t}\frac{1+\beta}{\beta}k_{t+1}. \tag{3.13}$$

The proportionality between the two types of capital results from the fact that wages provide a common source of funding for both types of capital; i.e. wages determine both private saving and the government's tax base. Think of the tax rate as being constant over time to simplify thing a bit. Using (3.4b) and (3.13), dated for period t, allows us to write (3.12) as the following transition equation for private capital,

$$k_{t+1} = \kappa k_t^{\alpha+\mu(1-\alpha)}, \tag{3.14}$$

where $\kappa \equiv (1-\alpha)A\left[\frac{\beta(1-\tau)}{1+\beta}\right]^{1-\mu(1-\alpha)}[\tau]^{\mu(1-\alpha)}$. The steady state value of k_t is $\bar{k} = \kappa^{\frac{1}{1-\alpha-\mu(1-\alpha)}}$.

There are two important differences between (3.14) and previous transition equations. First, the exponent on k_t has increased. The larger exponent makes the transition equation less concave and reduces the growth slowdown as capital accumulates—use a sketch of a transition equation that is less flat than in Chap. 2 to verify this (also see *Problem* 12). Recall from Sec. 3.2.4, that the sharply diminishing output growth rates and interest rates were problematic predictions of the neoclassical growth model. The fact that public capital rises with private capital reduces this problem, helping to smooth growth rates over the transition to the steady state and thereby providing a better fit to the historical experience of developing economies. The intuition is that as private and public capital move together, the rise in government capital raises the marginal product of private capital, reducing the force of diminishing returns.

Second, the leading coefficient of the transition equation, κ, is a function of the tax rate. One can show that the tax rate that maximizes the coefficient, and therefore the height of the transition equation, is $\tau* = \mu(1-\alpha)$. Tax rates above or below $\tau*$ will fail to maximize growth in private capital intensity. Tax rates that are too low do not generate enough public capital and tax rates that are too high cost too much in reduced private saving. The tax rate $\tau*$ just balances these opposing effects to make private capital as large as possible. The connection between tax rates and the growth in the private capital stock suggests that the size of government can be too small or too big from the perspective of maximizing a country's growth rate. See *Problems* 3 and 6–8 for more details. The chapter Appendix and *Problems* 9–11 discuss the best tax rates for achieving other objectives.

3.3 Public Capital and Productivity

We assume that public capital raises worker productivity, i.e. that $\mu > 0$. There is an empirical literature that attempts to test this assumption. The concept of public capital is quite broad and can include physical infrastructure, the stock of basic research knowledge, human capital acquired via public schooling, and even the intangible capital reflected in a country's laws and regulations—including the rules and procedures for implementing them. Empirical studies typically use national income accounting measures of public capital that are limited to physical infrastructure. Although there is some debate over the exact estimate of μ, most studies finds a positive and statistically significant effect of public infrastructure on output.

The classic empirical study of the productivity effects of public infrastructure was conducted by David Ashauer (1989). His approach allowed for a direct measure of μ, the output elasticity of public capital, which he estimated to be as high as 0.40. Subsequent research attempted to verify his findings, using different data sets and econometric approaches, and found a somewhat lower elasticity. Glomm and Ravikumar (1997) survey the empirical work in the decade following Ashauer's study and conclude that a more reasonable estimate might be 0.20. In an update of his earlier study, Ashauer (2000) found estimates close to 0.30. Several more recent studies also find estimates clustering around 0.30 (see the survey in Bivens (2012)).

It would be useful to have estimates of the effects that extend beyond public physical infrastructure. Less tangible types of public capital may have output elasticities that differ from physical infrastructure. Ideally one would decompose public capital into its different components. For example, a recent study has estimated a parameter very similar to μ that measures the human capital elasticity of public school spending. Interestingly, Manuelli and Seshadri (2014) find a public school spending elasticity estimate of about 0.30. Their estimate is based on an assumption that public school spending has a rate of return similar to that of private physical capital, about 7%. Heckman and others argue that, at the levels of school spending seen in developed countries, the marginal rate of return to public school spending in the average community is much lower than 7% (Heckman and Krueger (2005)). This is consistent with the historical analysis of Rangazas (2000, 2002) who finds a public spending elasticity of less than 0.20.

Another measurement issue in empirical studies is related to the quality of public capital and government corruption (Chakraborty and Dabla-Norris (2011)). As discussed in the introduction, large portions of the funds officially budgeted for public investment are never actually invested but instead are siphoned off for consumption by public officials and private contractors. In addition, the effectiveness of the public capital that does exist is influenced by how it is maintained and operated by government bureaucrats. This issue not only applies to infrastructure, power plants, and water and sewage facilities, but also to public schools where teacher absenteeism is a problem. The inability to control for these measurement issues will create a downward bias in the estimates of output effects from public capital.

3.4 Pure and Impure Public Capital

Thus far we have assumed that public capital is a private good, similar to private capital. With private capital, if one worker drives a tractor or operates a computer, then it is not possible for another worker to use the same equipment to produce output. For some types of public capital, the analogy to private capital is not accurate. If a producer is using a public road, this may not inhibit another producer from using the same road, at the same time, in any significant way.

If the transportation services provided by the road are not affected by the total number of producers using the road, then the road would be a *pure public good*—no "crowding" or reduction of services occurs as the number of producers served increases. Roads, while not pure private goods, are not pure public goods either because when the road becomes sufficiently busy with traffic, the total number of producers using the road *does reduce* the transportation services provided per producer. Roads, and many other types of public capital, are best viewed as *impure public goods* where crowding can occur.

This discussion affects the modelling of the production function that relates public capital to output. If public capital were a pure public good, instead of a private good as in (3.3), we would write the labor productivity index as

$$D_t = E_t^{1-\mu} G_t^{\mu} \tag{3.15}$$

where now the *total* public capital stock determines the productivity of an *individual* producer, independent of how many producers there are in the economy.

A more general way of writing the productivity, that includes pure private and pure public goods as special cases, introduces *impure public goods*,

$$D_t = E_t^{1-\mu} \left({}^{G_t}/_{(N_t)^{\xi}} \right)^{\mu} \tag{3.16}$$

with $0 \le \xi \le 1$. The parameter ξ gauges the public goods nature of public capital. If $\xi = 1$, then public capital is a private good, as in the case of private capital. If $\xi = 0$, then public capital is a pure public good. For $0 < \xi < 1$, we have an impure public good, where some crowding occurs.

Now we need to think about how taking the simple route of modeling public capital as a private good, when in fact it is more accurate to model it as a impure public good, affects the analysis. Toward this end, note that we can write (3.16) as

$$D_t = E_t^{1-\mu} \left(\frac{G_t \;\; N_t}{(N_t)^{\xi} N_t} \right)^{\mu} = \bar{E}_t g_t^{\mu} \tag{3.17}$$

where $\bar{E}_t = E_t (N_t)^{(1-\xi)\mu}$. The general productivity index in (3.17) has the same form as (3.3), but with an adjusted expression for exogenous technological change. This means, even if public capital is an impure public good, we can continue to model it as a private good. However, the adjusted technological progress will increase with population size. For a given ratio of public capital per producer, a larger economy will generate more output *per producer*. This is because the producers, at least to some extent, can share the total public capital, and with more producers there is a greater total public capital stock for any given value of g_2. Note that the *sharing effect* diminishes with population size because $(1-\xi)\mu < 1$. So, for large populations, variations in population size do not affect worker productivity very much, when g_2 is held constant.

The lesson here is that we can model public capital as a private good and use (3.3), but we have to remember that the technology index is a function of population size if public capital has public good characteristics. For most of our analysis, this consideration will not be important.

3.5 Capital Accumulation in an Open Economy

We have been working under the assumption that the economy is perfectly closed to international trade. Suppose now that private capital owners have the option of investing their capital across borders. The simplest way to introduce an international market for funds is to assume a sufficiently large number of countries are trading with each other. When many countries are engaged in trade, it may be reasonable to assume that the international market for funds is perfectly competitive at the level of an entire country. A single country is so small relative to the entire market that they take the international interest rate as an exogenous variable beyond its influence. This assumption is most accurate for smaller economies, so an open economy model with an exogenous international interest rate for funds is called the *small open economy* model.

Assume that the domestic capital owners reside in a small open economy. The perfectly competitive international rental rate on physical capital is an exogenous variable denoted by r^*. The capital owner's return to investing capital in foreign countries is then $1-\delta+r^*$. If instead the capital is invested domestically, the return is $1-\delta+\alpha A K_t^{\alpha-1}(D_t N_t)^{1-\alpha}$. In equilibrium these two returns must be the same.

Taking the same steps followed in (3.4a) to write the domestic marginal product expression in capital intensive form, the open economy equilibrium requires that $r^* = \alpha A g_t^{\mu(1-\alpha)} k_t^{\alpha-1}$. We can then solve for the open economy domestic capital intensity as

$$k_t = \left[\frac{\alpha A g_t^{\mu(1-\alpha)}}{r^*}\right]^{\frac{1}{1-\alpha}}. \tag{3.18}$$

Notice that there are no variables related to the domestic country's national saving (such as β or wage taxes). When foreign direct investment is possible, national saving does not affect capital intensity. If the fundamentals determining the return on capital are attractive (high values for A and g), foreign saving will flow into the country to build up the domestic capital stock. This suggests that a country may want to focus its policies more on the foundations of a high marginal product of capital than on increasing national saving. For example, in an open economy a wage tax that finances public infrastructure is particularly appealing because the drop in private saving will not drive up domestic interest rates and crowd out private investment—in fact the improved public infrastructure will unambiguously attract private capital from abroad.

The problem with this strategy is that a low saving country can become quite dependent on the conditions in international loan markets. If high saving countries begin supplying fewer funds to international markets, r^* will increase, causing a fall in the k of a low saving country. This is precisely a danger for the United States. Fiscal policies in the United States have reduced national saving and increased dependence on foreign saving (see Ivanyna et al. (2018)). Several forces suggest that shortages of international funds will develop in the future that may significantly raise r^*, hurting international borrowers such as the United States.

3.5.1 Open Capital Markets and Growth in Developing Countries

While capital scarcity should attract funding from abroad, the empirical evidence supporting the connection between open capital markets and economic growth is inconclusive (see, for example, Kose et al. (2009)). One reason an open capital market might not attract foreign funding for investment in a developing country is that an unusually low capital-labor ratio does not necessarily imply an unusually high return to investment.

To see this point explicitly, note from (3.4a) and (3.18) that the marginal product of capital is not only a function of k but is also a function of A and g. If a capital scarce country also has low levels of TFP or public infrastructure, the domestic marginal product of capital could be *lower* than the equilibrium return to capital in global capital markets. To generate the high return that attracts foreign capital, a country must have policies that support adequate levels of human and public capital.

The inconclusive empirical findings have inspired more thinking about why the growth effects of open capital markets have been difficult to identify. Recent research has focused on (i) new mechanisms through which growth may be indirectly promoted from openness (ii) a more detailed examination of the different forms of foreign investment and (iii) pre-conditions that a country might need in order to benefit from foreign investment (similar to the discussion in the preceding paragraph).

One type of indirect mechanism that has been considered is the connection between opening capital markets and domestic policies. In particular, some argue that the decision to open a country's capital market can act to discipline the country's monetary, fiscal, and regulatory policies to be more "pro-growth," so that the country can successfully compete for international capital. For example, Chap. 5 discuss this possibility using an overlapping generations growth model very similar to the one used in this chapter. A key difference is that the Chap. 5 model includes an *endogenous* theory of fiscal policy. We show that the optimal fiscal policy changes when an economy opens its capital market. In the open economy, private capital formation is more responsive to tax rates and public capital. This creates an incentive to lower tax rates and to increase the share of a given budget that it devoted to public investment, which not only attracts foreign capital but also increases growth directly.

Foreign investment can be decomposed into *portfolio* investment—financial capital supplied when foreign investors purchase domestic stocks, bonds, and bank accounts, and *direct* investment—physical capital that is financed and managed by foreign multinational firms. Recent findings suggest that opening equity markets increases economic growth, while the growth effects from opening bond markets and from foreign direct investment (FDI) are less clear (see, Kose et al. (2009)).

The lack of clear growth effects from FDI is particularly puzzling. Economists have traditionally believed that FDI is more beneficial to a developing countries growth than portfolio investment for two reasons. First, in addition to augmenting the domestic capital stock, FDI may have effects on the domestic country's TFP through transfers of technology and managerial practices. Second, FDI is harder to suddenly reverse, making it less volatile than inflows of financial capital. So, why aren't there clear growth effects from FDI?

Alfaro (2016) provides a survey of the recent attempts to answer the question. In many countries FDI does not actually bring its own financing. Often foreign companies attempt to finance physical capital formation by raising the funds in the destination country—which has the potential to reduce funding for domestic firms. It also appears that for the domestic country to benefit from technological spillovers, certain preconditions must be met. The domestic economy must have threshold levels of human capital and reasonably developed financial markets for workers and domestic firms to benefit from and replicate the new production methods tied to FDI.

The main overall lesson is that a developing country can accelerate growth by opening its capital markets, but only if its domestic policies have laid the foundation for high returns to private capital—a literate and numerate workforce, reliable public infrastructure, and the beginnings of a financial sector.

3.6 Government: Benevolent Dictator or Kleptocrat?

Previous sections treated fiscal policy as exogenous. This section develops a theory of *why* the government chooses particular policies—an introduction to *political economy*. We begin with the same private sector structure as before; the overlapping-generations growth model is briefly summarized here. Then we introduce a theory of fiscal policy formation that includes both selfish and altruistic concerns of the government.

3.6.1 Firms

Production takes place within standard neoclassical firms that combine physical capital and human capital to produce output from a Cobb-Douglas technology

$$Y_t = AK_t^{\alpha}(D_t N_t)^{1-\alpha}. \tag{3.19}$$

The productivity index, D, is now a function of disembodied technology, E, and government capital per adult worker, G/N, and is given by

$$D_t = E_t^{1-\mu}(G_t/N_t)^{\mu}, \tag{3.20}$$

where $0 < \mu < 1$ is a constant parameter (ignore the crowding of G by public officials for simplicity). This specification captures the idea that public infrastructure raises the productivity of the private sector. We assume that E progresses at the exogenous rate q and the exogenous growth factor of the population is n.

Firms operate in perfectly competitive factor and output markets. They choose physical capital (K_t) and effective labor ($H_t = D_t N_t$) to maximize profit. The profit-maximizing factor mix must satisfy

$$r_t = \alpha A g_t^{\mu(1-\alpha)} k_t^{\alpha-1} \tag{3.21a}$$

$$w_t = (1-\alpha)A g_t^{-\alpha\mu} k_t^{\alpha}, \tag{3.21b}$$

where the de-trended, for exogenous technical progress and population growth, values of public and private physical capital are defined as $g \equiv G/EN$, and $k \equiv K/EN$. The wage paid to a worker, with embodied skills indexed by D_t, is $w_t D_t = (1-\alpha)AE_t g_t^{\mu(1-\alpha)} k_t^{\alpha}$. Note also that $y_t = AE_t g_t^{\mu(1-\alpha)} k_t^{\alpha}$, where $y_t = Y_t/N$, output per worker.

3.6.2 Households

Households maximize the utility function $U_t = \ln c_{1t} + \beta \ln c_{2t+1}$ subject to the lifetime budget constraint, $c_{1t} + c_{2t+1}/R_t = (1-\tau_t)w_t D_t$, where $R_t \equiv 1 + (1-\tau_t) r_{t+1} - \delta$ and τ_t is the proportional net tax rate on income. As before, we assume $\delta = 1$, so $R_t = (1-\tau_t)r_{t+1}$. The resulting optimal consumption and saving behavior is given by

$$c_{1t} = \frac{1}{1+\beta}(1-\tau_t)w_t D_t \tag{3.22a}$$

$$c_{2t+1} = \frac{\beta}{1+\beta}R_t(1-\tau_t)w_t D_t \tag{3.22b}$$

$$s_t = \frac{\beta}{1+\beta}(1-\tau_t)w_t D_t. \tag{3.22c}$$

3.6.3 Capital Market Equilibrium

The firm's demand for private capital intensity is implicitly given by the profit maximizing conditions in (3.21a and 3.21b). The supply of private physical capital from households is made available for firms to rent in the factor markets and is given by

$$k_{t+1} = s_t N_t. \tag{3.23}$$

Using (3.21b, 3.22 and 3.23), the equilibrium transition equation for physical-capital intensity is

$$k_{t+1} = \frac{\beta}{1+\beta} \frac{(1-\tau_t)(1-\alpha)}{(1+q)n} A g_t^{\mu(1-\alpha)} k_t^{\alpha}. \tag{3.24}$$

3.6.4 Government

We now introduce a "reduced-form" approach to the formation of fiscal policy. The government is run by public officials that are distinct from private households in that they derive their income from public funds and set fiscal policy with the entire future path of the economy in mind. This specification is similar to the common approach in macroeconomics of modeling the government as a benevolent social planner. Here we extend that approach by letting the degree of government altruism vary. There is no deep model of the politics that determine how the government is chosen and how their policies are influenced by voters and interest groups. Instead we take as given the politics of a country that determine the "reduced-form" preference parameters of the government. The parameters dictate the government's concern about the welfare of the general population of private households and the welfare of households that make up, or are closely connected to, the government itself. The deeper political determinants of these reduced form parameters are assumed to be given throughout the analysis. Thus, we examine how policies are formed within a given political environment.

In short, we do not believe that there is a unique mapping from political institutions to the government's preferences over economic policies. Pro-growth policies may be carried out and implemented within a highly democratic political process or by a completely authoritative dictator (think of the dictators that pushed development during the Asian Tiger "Growth Miracles"). Different political institutions can give rise to similar reduced-form preferences of the policy maker. In addition, attempting to model the politics of a country is complex and requires that compromises be made in the economic modeling. Jointly modeling political and economic equilibria is particularly difficult in the economic environments that we focus on in this book—the transitional growth of overlapping generation economies.

We assume the government officials who determine fiscal policy are some fraction, ς, of the population of private households, N_t. Government officials value their own consumption () as well as the welfare of the representative citizen according to a single period utility function, $\ln c_t^g + \phi U_t$, where ϕ is a positive preference parameter that gauges the relative weight the government places on the welfare of private households, U_t [1] We assume the current government also cares about the government as an on-going institution (i.e. they care about the future operations of the government and the welfare of future government officials) and the welfare of the country's future citizens. The preferences of the government are given by[2]

$$\sum_{t=0}^{\infty} \beta^t \left(\ln c_t^g + \phi U_t\right). \tag{3.25}$$

These complicated preferences make explicit that the government's concerns extend indefinitely into the future. This is because there is no natural time horizon for government planning. Maximizing an objective function such as (3.25) is somewhat difficult but it turns out that the solutions for the optimal fiscal policy are surprisingly simple.

The government budget constraint is

$$c_t^g \varsigma N_t = \tau_t Y_t - G_{t+1}. \tag{3.26}$$

The left-hand side gives the government's consumption expenditures. The right-hand side is the difference between government tax revenue, net of transfers, and government expenditures on public capital. For simplicity, we assume that both private and public capital fully depreciates over what we assume to be 20–30 year-long periods of the model. Next period's public capital stock is determined solely by this period's public investment.

To find the optimal fiscal policy, the government chooses sequences of tax rates, government consumption, and government capital to maximize the discounted utility of government officials and private households, given by (3.25), subject to a series of the budget constraints and private capital accumulation equations given above.[3] In addition, the government takes into account how their policy choices affect all private sector decisions. The solution to the government's problem is[4]

---

[1] Mulligan and Tsui (2015) present a theory, based on the threat of political entry, that can be viewed as making γ endogenous.

[2] For notational simplicity only, we assume the government's time discount factor is the same as that used by private households. One could allow the discount factor to differ from private households to study how the government's time preference affects policy.

[3] We assume that the government can commit to its policy choices in advance. For a discussion of commitment issues in regard to the setting of fiscal policy see Lundquist and Sargent (2004, Chapter 22).

[4] See Chap. 5 for a sketch of the derivation in a somewhat more complicated economy that includes the current model as a special case.

$$\tau_t = \tau = \frac{1 - \alpha\beta + \beta\mu(1-\alpha)2\phi}{1 + 2\phi}, \tag{3.27a}$$

$$g_{t+1} = \frac{\beta\mu(1-\alpha)}{(1+q)n} A k_t^{\alpha} g_t^{\mu(1-\alpha)}, \tag{3.27b}$$

$$k_{t+1} = \frac{\beta(1-\tau)(1-\alpha)}{(1+\beta)(1+q)n} A k_t^{\alpha} g_t^{\mu(1-\alpha)}. \tag{3.27c}$$

Equation (3.27a) tells us the tax rate is constant over time. One can show that the constant tax rate τ is decreasing in ϕ, more concern for private households implies a lower tax rate. Equation (3.27b) gives a transition equation for the public capital stock that is analogous to that for the private capital stock. Here, the government's saving rate out of national income is a constant, $\beta\mu\,(1-\alpha)$. Combined with (3.27a) this tells us that a more selfish government, with a lower ϕ, will collect more in taxes but invest *a smaller fraction of tax revenue* in public capital so as to maintain the *same* investment rate out of national income. Equation (3.27c) simply repeats the transition equation for private capital accumulation.

Note that, as in Sect. 3.1, we can use (3.27b) and (3.27c) to reduce the dynamics to that based only on the private capital-labor ratio

$$k_{t+1} = \kappa k_t^{\alpha+\mu(1-\alpha)}, \tag{3.28}$$

where $\kappa \equiv \frac{\beta(1-\alpha)A}{(1+q)n}\mu^{\mu(1-\alpha)}\left(\frac{1-\tau}{1+\beta}\right)^{1-\mu(1-\alpha)}$.

3.6.5 Steady State Equilibria and Income Gaps

Using (3.27b) and (3.27c), the steady state equilibrium is characterized by the following expressions for the private and public capital intensities,

$$\bar{g} = \frac{\mu(1+\beta)}{1-\tau}\bar{k}. \tag{3.29a}$$

$$\bar{k} = \kappa^{\frac{1}{1-\alpha-\mu(1-\alpha)}}, \tag{3.29b}$$

which implies

$$\bar{y}_t = AE_t\Omega(1-\tau)^{\frac{\alpha}{1-\alpha-\mu(1-\alpha)}}, \tag{3.29c}$$

where $\Omega \equiv \left([\mu(1+\beta)]^{\mu(1-\alpha)}\left[\frac{\beta}{1+\beta}\frac{1-\alpha}{(1+q)n}A\right]^{\alpha+\mu(1-\alpha)}\right)^{\frac{1}{1-\alpha-\mu(1-\alpha)}}$.

Using (3.29), we can compute differences in worker productivity due solely to differences in fiscal policy (based on a differences in ϕ that work through τ). We think of a low-tax "rich" country (R), with a government that behaves like a benevolent dictator, and a high-tax "poor" country (P), with a government that behaves like a kleptocrat. The steady state income ratio for these two countries is

$$\frac{y^R}{y^P} = \left(\frac{1-\tau^R}{1-\tau^P}\right)^{\frac{\alpha}{1-\alpha-\mu(1-\alpha)}}. \tag{3.30}$$

Fiscal policy is by no means the primary reason why incomes differ across countries, as we demonstrate below. However, it is a reasonable candidate because there are several poor countries with unusually large governments.

Table 3.1 gives examples of poor countries with levels of τ, or government purchase shares, that are about double those of the US. The average government purchase share of those countries is 0.32. The U.S. purchase share is typically between 0.15 and 0.20. The comparison is for 1985, a year that generates close to the largest income gaps between the U.S. and most the African countries during the twentieth century. Starting in the 1990s, Africa began growing faster. Most of the countries in Table 3.1 have grown between 4 and 9% per year since the mid-1990s. The exceptions are the Central African Republic and Comoros, whose growth rates remain low and thus have seen their income gaps expand.

To quantify the model's predictions about income differences due to fiscal policy, we need to calibrate the model's parameters. The physical capital income share,α, is set to the standard value of 1/3. Based on the review of the empirical literature in Sect. 3.3, the output elasticity for public capital, which here is $\mu(1-\alpha)$, is set to 1/3.

Forming an extreme case from the data above, we set the rich country tax rate at 0.15 and the poor country tax rate at 0.35. The gap in the tax rates causes a gap in income of about 30%. While this is a significant difference in income, it does not come close to explaining the huge differences seen in Table 3.1. Chapter 5 extends the model of fiscal policy differences to also include human capital and fertility differences from Chap. 4. These extensions are able to generate the large income

Table 3.1 Government size—selected low-income countries (1985)

Country	Government purchases/GDP	$y^{US}/y^{country}$
Angola	0.36	11
Burkina Faso	0.29	33
Central African Republic	0.44	17
Comoros	0.49	10
Ethiopia	0.28	40
Gambia	0.37	17
Mozambique	0.31	33
Uganda	0.28	33
Average	**0.32**	**24**

Source: Alan Heston, Robert Summers, and Bettina Aten, Penn World Table Version 6.1, Center for International Comparisons at the University of Pennsylvania, October 2002 gaps observed in Table 3.1. Human capital differences not only directly affect worker productivity differences, but also indirectly create private (via saving) and public (via the tax base) physical capital differences.

Apart from the government's important role in promoting human capital formation, one can think of other ways the government's contribution to explaining cross-country income differences might be expanded. First, the estimates of α and μ, that determine the quantitative impact of fiscal policy differences on income gaps, may be too low because they are based strictly on measures of tangible capital. As emphasized by Parente and Prescott (2000), in the case of private capital, there are substantial investments in building *intangible* capital. Private firms make investments in research and development of products and production techniques as well as in the specific human capital of their work force. The same considerations could be applied to government investment in improving laws, regulations, and the efficiency of bureaucracies. Expanded notions of capital can be used to motivate larger estimates of α and μ, and thus larger income gaps due to tax differences across countries.

Second, the effectiveness of pubic capital may differ across rich and poor countries. For example, there is evidence suggesting that less than half of the funds in public capital budgets are actually invested in some countries (Ivanyna et al. (2018)). We can capture this possibility here in a simple way by writing a new productivity index as $\tilde{D}_t = E_t^{1-\mu}((1-u)G_t/N_t)^{\mu}$, where u is a parameter that takes values between zero and one, representing the fraction of the investment budget that is diverted toward public officials and private contractors. In developing countries, because of low-quality governance, the value of u may be high relative to rich countries with more checks on corruption or more experience in managing public investment projects. We can pull $1 - u$ out of the expression for D and write the production function as $Y_t = \tilde{A}K_t^{\alpha}(D_t N_t)^{1-\alpha}$, where $\tilde{A} \equiv (1-u)^{\mu(1-\alpha)}A$. Assuming that $u = 0$ for the rich country, overly optimistic for sure, we can rewrite (30) as

$$\frac{y^R}{y^P} = \left(\frac{1}{1-u}\right)^{\frac{\mu}{1-\mu}} \left(\frac{1-\tau^R}{1-\tau^P}\right)^{\frac{\alpha}{1-\alpha-\mu(1-\alpha)}}. \tag{3.30$'$}$$

So now the income gap depends on *two* aspects of fiscal policy, u and τ, the effects of which depend on the values for α and μ. The end-of-chapter *Problems* will explore these extensions further. Ivanyna et al. (2018) offer a much more complete analysis of corruption, including features that mediate its negative impact on economic growth.

3.6.6 Severe Government Failure

Poor government policy and corruption prevent economies from reaching their potential living standards. However, some countries' economic performance is so abysmal that the direct effects of poor economic policies alone cannot explain their plight.

The post WWII period has seen a wide variety of long-run growth experiences as reflected in the stylized growth fact *G6*. From 1960 to 2000, nine Growth Miracle countries generated annual growth rates in per income of more than 4%, led by Singapore's remarkable growth rate of over 7% (Weil (2005, Figure 1.6)). Equally surprising are 26 Growth Disaster countries, whose initial incomes were quite similar to the Growth Miracle countries in 1960, that saw their per capita income *fall* over the four decades. Their living standards were worse in 2000 than in 1960!

An important factor explaining growth disasters is political violence within the country, including frequent strikes and protests, military coups, and civil wars. Political violence is the dramatic outcome of a government that has failed to establish a reasonable standard of living and a fair distribution of income across households.

Paul Collier's work has demonstrated that a "conflict trap" can prevent countries from sustaining the modern growth needed to escape poverty (Collier 2009). He and his coauthors show that low income and slow economic growth increase the probability of civil war. In poor countries with substantial ethnic diversity and dependence on income from natural resources, the probability of internal conflict is particularly high. In these settings there is fierce competition for the rents from the natural resources and little attention paid to policies that would raise living standards for all groups. Even if a war does not breakout that would obviously destroy an economy, the insecurity associated with a high probability of conflict discourages investment in physical and human capital and reduces growth.

Collier believes a country stuck in a conflict trap needs help from external sources. Neighboring countries, who also suffer harm due to spillover violence, should have the incentive to help the country in conflict maintain political stability. With the support of the broader international community, Collier recommends an external coalition of peacekeepers to provide security to governments that are established by elections meeting the international standards of fairness. In hopes of raising government accountability to the country's population, Collier essentially proposes a trade of internally provided security for fair elections. He also recommends that financial aid to a troubled country be contingent on the country (i) receiving technical assistance in establishing a transparent accounting system that tracks government spending and (ii) cutting military spending (that has been shown to increase the probability of civil war).

3.7 Slowing Long-Run Economic Growth

After adding human capital, the historical simulation from Sect. 2.5 of Chap. 2 displays no clear downward trend in growth rates from 1870 to 1990. The model's ability to explain trendless growth is further enhanced by adding public capital, comprised of a variety of inputs that are complementary to private physical capital. However, even if investment rates are maintained, the fact that $\alpha < 1$, $\mu < 1$, and $\theta_1 + \theta_2 < 1$ implies that diminishing return must eventually dominate and cause a growth slowdown—one of the key predictions of neoclassical growth theory.

Is there any evidence of a growth slowdown? After more than a century of trendless growth, there is now enough data to detect that growth is, in fact, slowing.

It appears that worker productivity growth in the U.S. began to permanently slow in the 1970s.[5] From 1920 to 1970, the annual growth rate in worker productivity was 2.82%. Since 1970, the growth rate has been more than a full percentage-point lower at 1.62%. The OECD countries as a whole saw very high growth in worker productivity during the recovery from World War II, with an average annual growth rate of 4.3% from 1950 to 1972.[6] From 1972 to 1995, the growth rate naturally slowed from the high post-war recovery rate down to 2.4%. Since 1995, growth rates have fallen further, down to just 1.4%.

The robust economic growth after World War II allowed government expansion because it brought with it large increases in tax *revenues* without the need to raise tax *rates*. Even as economic growth rates began to slow, and budget deficits began to appear, it was natural for politicians and citizens to believe that the relatively high growth rates, seen for decades after WWII, would return. There was, and still is, optimism that computer-related technological advances would raise economic growth rates above than those of the twentieth century, helping to rescue us from our fiscal problems.[7] However, the computer-related technological advances have been with us for some time, including the 45-year period over which economic growth rates have fallen considerably. Computer-driven technological progress has not stimulated economic growth the way earlier twentieth century technological advances did.[8] To maintain the current relatively modest growth, new sources of growth will have to be quite dramatic to offset other forces that continue put downward pressure on growth rates around the developed world. There are three forces, in particular, that will continue to pull growth rates down over this century unless we change our policies.

3.7.1 Reduced Saving and Investment

The first force pulling economic growth rates down is the decline in domestic saving and investment. Economic theory predicts intergenerational transfers from younger and future generations to older generations, in the form of government retirement programs, raise consumption and lower saving (Ivanyna et al. (2018, Chapter 4)). Consistent with this prediction, the U.S. has seen a decline in its net national saving rate. The net national saving rate averaged about 15% of GDP from 1950 to 1975.[9] Since then it has declined significantly. Even before the Great Recession, the net national saving rate was below 4%. Similar trends are present in other economies, as

[5]Gordon (2016, Figure 1–2).

[6]OECD (2015, Table A1).

[7]See, for example, Brynjolfsson and McAfee (2014).

[8]See Viig (2011) and Gordon (2016).

[9]Kotlikoff (2015, Chart 2).

saving has fallen across the developed world. Just as in the United States, the decline in saving is associated with societies placing increasing weight on current consumption, which is partly reflected in greater intergenerational transfers toward older households.[10]

In the U.S., domestic investment has not declined to the same extent as national saving because of an influx of foreign saving. Most of the foreign funding in U.S. financial markets has come from Japan and China. However, Japan has its own fiscal crisis and China is seeking to expand its domestic consumption rate. Thus, the continued supply of foreign funding is in serious question. The scarcity of international funds will also be affected by the fact that many other developed countries will be seeking foreign financing for their expanding public debt. This all means that domestic investment is soon likely to fall more closely in line with national saving.

The rise in government funding for consumption of retired households has also been associated with a decline in *public* investment—a main focus of this chapter. Government infrastructure investment in the U.S. measured about 3.5% of GDP in 1970. Today, it is about 2.5%. Net investment, after accounting for depreciation, is currently only about 0.5% of GDP. As a result of this decline, the public infrastructure of the United States has depreciated to an embarrassing state for such a rich country.[11] Public infrastructure investment has been neglected in other developed countries as well.[12] Another important public investment is the financing of basic research. The fraction of federal funding for the basic research, that lays the foundation for technological progress, was cut over the last quarter of the twentieth century.[13] In addition to budget pressures that are crowding out public investment, there is the concern that individual governments now face reduced incentives to invest in basic research because of the increased ease of international spillovers of knowledge.[14] If each country attempts to free ride off the basic research of other countries, technological progress across the globe will fall.

3.7.2 Slowdown in Human Capital Growth

The second negative force on growth is the decline in human capital accumulation. The slowdown in human capital formation has occurred along several margins—years of schooling, skill acquisition within a school-year, and pre-school investments in young children.

[10]Dobrescu et al. (2012).

[11]See Friedman and Mandlebaum (2012), Malinovskaya and Wessel (2017), and Stupak (2018).

[12]Aghion et al. (2013).

[13]Viig (2011).

[14]Viig (2011).

The average years of schooling across OECD countries increased from 10 to 12 between 1990 and 2013. It is predicted that it will take more than 50 years for the average to increase from 12 to 14.[15] The slowdown in the growth of years of education has been more dramatic in the United States, which has lost its position as the most educated country in the world. The age-cohort born in 1925 received 10.9 years of schooling, while those born 25 years later in 1950 received 13.2 years, a gain of 2.3 years. Moving forward another 25 years, saw those born in 1975 receive 13.9 years of schooling, a gain of only 0.7 years.[16]

The slowdown in the growth of years of schooling is due to the inability of rich countries to significantly raise their college enrollment rates. In the United States, the four-year college participation rate for high school graduates, age 23 and under, has shown little trend since 1970—never consistently rising above 60%.[17] College completion rates by age 23 have also been trendless at less than 20% of the age cohort.

The modest rise in years of schooling has been, in part, due to a rise in the enrollment and completion rates for older students.[18] By age 30 about 30% of the age-cohort obtains a 4-year degree, a little less than half of those who initially enroll. Some of the rise in years of schooling is also due to more 18–19 year olds enrolling into 2 year colleges after high school. The percent of 18–19 year olds, who have completed high school and are enrolled in some type of college has risen from 60% in 1990 to 66% in 2013.[19] The rise was almost entirely due to increased enrollment in 2-year colleges. The percentage enrolled in 4-year colleges was essentially flat at 40%. For 2014 and 2015, enrollments rates for 4-year colleges have continued to be flat, while enrollment rates in 2-year colleges have actually fallen.[20]

The modest rise in the years of schooling overstates the rise in human capital because all indicators suggest a decline in skills acquired by the average college student. The data we have on the quality of education is for the United States, but quality issues may be an explanation for the slowing growth in years of schooling across the OECD countries generally.

The record of college preparedness in the United States is particularly poor for such a rich and highly educated country. On the Program for International Student Assessment (PISA) test, taken by 15 year-olds across 34 OECD countries, the United States ranks 27 in math, 20 in science and 17 in reading. The relatively poor performance of the United States on the PISA test has not changed over time. Despite rising real expenditures on high school students, national test scores have also been relatively flat for the past 50 years. In fact, the test scores have recently

[15]OECD (2014, Figure 2).

[16]Gordon (2016, p.513) and Katz (2005, pp.270–274)).

[17]Carneiro and Heckman (2005, Figure 2.2 (a)) and Turner (2004, Figures 1.1, 1.2).

[18]Turner (2004, Figure 1.5).

[19]National Center for Education Statistics (2015).

[20]NSC Research Center (2015).

dipped and hit lows that haven't been seen for decades.[21] Performance on measures of adult skills (basic literacy and numeracy needed for work) has also fallen off. OECD measures of basic skills peaked for cohorts born between 1978 and 1987 and have fallen since. The recent decline in scores is largest for the United States. Only about 40% of high school graduates are deemed prepared for success in college by their performance on the SAT and only 28% by their performance on the ACT. With at best a mediocre and stagnant track record in getting students ready for college, it is not surprising that enrollment and graduation rates are also relatively stagnant. Given that per pupil spending has risen over time at all levels of education, the obvious conclusion is that the marginal returns to human capital investments under current education policy are low.

Surprisingly, given the backdrop provided above, grades given in college courses are up. With no indication of an improvement in college-preparedness, the rise in grades suggests that standards and content in college are slipping and those who do graduate have less skills than in the past.[22] In 1960 about 33% of all grades given were As, today it is 43%. The rise in grades is even more dramatic at prestigious schools. In 1966, Harvard gave 22% As, in 2002 the percent of As was 46%. The rise in grades coincides with a *decline* in student study time. Students spend about 13 hours less studying today than in the 1960s.

The only explanation for the combination of flat college-preparedness, declining study time, and rising grades, is an elimination of course content and a lowering of standards. It is difficult to find older college professors who do not admit to eliminating content and lowering standards over their careers. In fact, it is becoming increasingly difficult to simply find a professor. In 1960, 75% of college instructors were full-time tenure track professors. Today the number is 27%.[23]

The labor market data for college graduates is also consistent with low or declining skills. Surveys of hiring managers have revealed that only 16% found college graduates well prepared with skills and knowledge needed for the job.[24] There has been a growing wage gap between college and high school graduates that seems to suggest that the market value of college students is increasing. This growing wage gap, however, is driven primarily by a relatively small fraction of students with *graduate* degrees. Recently, workers with only undergraduate degrees have been struggling to find good jobs and their average real wages have been falling over the last decade.[25] It is only the very highly educated that have seen their real wages rise significantly over the past 30–40 years.

While real wages for most college graduates are flat or even falling, the average *rate of return* to college for those that graduate has remained high. This is because

[21]See Adams (2016) and Hanushek (2005, pp.252–259) for U.S. SAT scores and Rothwell (2016) for OECD scores of basic workforce skills.

[22]Bennet and Wilezol (2013, Chapter 4).

[23]Bennet and Wilezol (2013, p.139).

[24]Bennet and Wilezol (2013, p.146).

[25]Abel and Deitz (2015).

the largest cost of college for most students is the opportunity costs of not working during the college years. The opportunity cost of college has been falling because the real wages of high school graduates have been falling for some time. If the majority of children in advanced countries are not going to graduate from college, as is apparently the case, then economic growth rates cannot be improved without raising the productivity and wages of those who do *not* attend college.

There is increasing concern about educational investment in young children from low-income environments, particularly in the United States, but in other advanced countries as well.[26] Raising the productivity of workers who do not attend college is a challenging task because trends in family structure and falling real incomes for less than highly educated workers are limiting opportunities for children. On the optimistic side, there is growing evidence of high returns to early investment in children from disadvantaged family backgrounds.[27] The fact that the returns to pre-school investment in children from low income families are higher than the returns to marginal public school spending in middle and upper class neighborhoods, suggests that a reallocation of public funding could increase growth and reduce inequality.[28]

3.7.3 Technological Progress to the Rescue?

A decline in growth rates due to the diminishing returns associated with physical and human capital accumulation is inevitable. History shows the negative effect on growth rates can be mediated temporarily by raising investment rates, especially in human capital as indicated by our analysis in Chap. 2. However, there are ultimately growth slowdowns as investment rates level off.

This scenario paints a pessimistic forecast for growth in the twenty-first century. One can become even more pessimistic if there are reasons to believe that technological progress cannot continue indefinitely at the same rate we observed in the twentieth century. Charles Jones (2002) relates technological progress to the growth in researchers (scientists and engineers engaged in research and development). In the twentieth century, the growth in researchers was based on population growth and on growth in research intensity (the fraction of the available work force devoted to research). Jones points out that the only growth that is sustainable comes from population growth (as with all investment rates, the fraction of the work force devoted to research is bounded). Assuming that population growth remains similar to that of the second half of the twentieth century, long-run growth is expected to be less than ½ percent.

[26]For the United States see Caneiro and Heckman (2005) and Putnam (2015). For the UK and the OECD as a group, see Aghion et al. (2013) and OECD (2014, p.45).

[27]Heckman et al. (2010).

[28]Caneiro and Heckman (2005).

The issue of twenty-first century growth was made popular by an article appearing in the *Economist* (January 12, 2013), entitled "Innovation Pessimism." The article presents another reason to be pessimistic about growth. Academic research suggests that there may also be diminishing returns to research and development efforts (which Jones (2002) does not assume). Recent research by Bloom et al. (2017) suggests there are clear diminishing returns to research effort. They conclude that larger and larger increases in research effort will be needed to maintain technological progress at its current pace. Vijg (2011) argues that the pace of technological progress will slow, and in fact has already begun to, particularly in the important areas of energy, transportation and medicine. This pessimism is contested by those who argue that the growth impact of innovations in computing, biotechnology, and personal communications has not yet been fully realized. Brynjolfsson and McAfee (2014) claim that we are just on the cusp of a second machine-age built around the computer and the development of artificial intelligence.

Another reason to suspect a decline in technological progress in developed countries relates to immigration patterns. Developed countries tend to attract high-skilled labor from developing countries. For example, survey studies by Vivek Wadhwa (2012) have revealed the importance of immigration for innovation in the U.S.. In the U.S. only 12% of the population is foreign born. However, this relatively small group has contributed about 25% of U.S. global patents. Foreigners, already in or looking to do business in the U.S., receive half of U.S. domestic patents. Immigrants are responsible for almost 30% of new business formation, an important determinant of job formation.

Econometric studies provide evidence consistent with the implications of Wadhwa's survey data. Hunt and Gauther-Loiselle (2010) estimate that a 1 percentage-point increase in the immigrant share of U.S. college graduates increases patents per capita by 9–18%. Their estimates suggest that over the 1990s, the 1.3 percentage point increase in the immigrant share of college graduates raised patenting per capita between 12 and 21%.

As is commonly known, many high-skilled immigrants are from China and India. Vivek sees evidence that high-skilled immigration from Asia into the U.S. is weakening. The reason is a combination of expanding opportunities in their rapidly growing home countries and the restrictions and delays associated with the U.S. visa process. Without reform of immigration policy needed to ease entry of high-skilled labor into the U.S., there will likely be a decline in innovation and entrepreneurial activity.

To maintain growth rates similar to the twentieth century, given the past importance of physical and human capital accumulation, it won't be enough to argue that technological progress will continue, it will have to accelerate. Given what we currently know, this seems unlikely. The Congressional Budget Office (CBO) computes the estimate of the fiscal gap by assuming that twenty-first century growth rates in worker productivity and per capita income will continue to be similar to what they have been in the late 20th and early twenty-first century, about 1.5%. If the gloomier growth rate predictions prove to be correct, the fiscal gap is actually larger than is currently estimated.

3.7.4 Summary

Beginning in the 1970s, growth rates exhibited a long-run downward trend in developing countries. Politics and economic fundamentals have created a pro-consumption bias in policy making. Intergenerational redistribution toward older households associated with fiscal policy has lowered national savings rates. The impact of lowered national saving on private investment has not yet been fully felt because of foreign funding of U.S. and European domestic investment by Japan and China, international saving flows that are not likely to maintain investment levels in the future. Government budget pressures created by the rising burden of financing consumption of the elderly have reduced spending on public infrastructure and basic research. Advances in years of schooling per worker have slowed because the fraction of the population attending and graduating from four-year colleges has weakly increased or stalled completely. Workers who are not highly educated have seen little or no increase in their productivity and real wages for decades.

3.8 Convergence

In *Question* 9 of Chap. 2, the concept of convergence was introduced. *Absolute convergence* is the idea that poor countries will generally grow faster than rich countries and "catch up" to their per capita income and living standards. The more qualified concept of *conditional convergence* argues that the catch-up only occurs if the poor country establishes similar fundamentals as the rich country, causing similar long-run steady states.

The neoclassical growth models used in Chap. 2 and in this chapter are consistent with conditional convergence. They predict that if the fundamentals of countries are similar then, regardless of when modern growth begins, all countries should eventually converge to similar long-run per capita income levels and living standards. Fundamentals refer to the preferences, technology, and policies of the country. The convergence prediction requires that countries beginning modern growth later must grow *faster* than countries that began modern growth earlier, in order to "catch-up."

In the standard neoclassical growth model of physical capital accumulation economic growth rates uniformly fall with the level of income as countries develop, making the prediction of convergence obvious and potentially rapid. However, the additions of human capital in Chap. 2 and public capital in this chapter suggest that convergence may be a very slow process. Investments in human capital and public capital can cause growth rates to exhibit little trend for many years. There could easily be no sign of convergence for decades after growth first begins. This may be true even if the fundamentals of the leading and lagging economies are similar.

Since World War II, most countries of the world have begun sustained economic growth. The standard neoclassical growth model of physical capital accumulation predicts that many of these countries, those that have established decent fundamentals, would grow faster than the leaders. However, the average growth rates of countries at every stage of development has been similar to that of the income leaders (see, for example, Jones and Vollrath (2013, Figure 3.6) and Kraay and McKenzie (2014, Table 3.1)). This means that, on average, we have not seen countries converge. The lack of convergence for many decades, while not consistent with the standard neoclassical model, is quite consistent with extended neoclassical model that include human capital and public capital. We revisit this topic in Chap. 9 in a two-sector model. There we show that over the course of development, growth rates first rise, then remain trendless for more than a century, before finally beginning to fall. This pattern is consistent with what we discussed in Sect. 3.7. In the developed world, after almost two centuries of trendless progress, growth rates have only fallen off in the last 40 to 50 years. Over these decades, investment rates in physical, human, and public capital have leveled off or declined. Under such a scenario, diminishing returns will dominate and growth rates will fall.

3.9 Exercises

Questions

1. Write down the most general version of the single period government budget constraint and explain what each variable represents.
2. Give a verbal description of how government capital is introduced into the growth model and how it affects worker productivity.
3. Describe the five determinants of private worker productivity, including the two new ones that relate to fiscal policy.
4. Explain how each of the following affect output per young household in the economy.
   (a) government purchases of private consumption goods
   (b) government employment
   (c) government investment
5. Use (3.14) to explain how the addition of government investments, such as public schooling and roads, can improve the simulation results from Chap. 2.
6. Summarize the attempts to empirically estimate the parameter μ.
7. If G_t is an impure or pure public good how does it change the analysis?
8. What are the main differences between how private capital accumulates in closed and open economies?
9. Use the profit-maximizing condition in an open economy, $r^* = \alpha A g_t^{\mu(1-\alpha)} k_t^{\alpha-1}$, to sketch the relationship between r^* and k_t. Your sketch should plot r^* and the domestic marginal product of capital on the vertical axis, and k_t on the horizontal axis. Use the sketch to explain why a country with low k_t, but also a low marginal product of capital curve, may actually lose capital, i.e. experience *capital flight*, if its capital markets are opened.

10. If a kelptocracy becomes more selfish, explain what happens to each of the following.
    (a) ϕ
    (b) τ
    (c) $\bar{k}$
    (d) $\bar{g}$
    (e) $\bar{y}_t$
11. What is the difference between a dictatorship and a kleptocracy? Can a dictatorship have a higher value of γ than a democratic government? Explain.
12. How are τ^R and τ^P calibrated? What does the calibrated difference in tax rates imply about the difference in worker productivity across countries? Does fiscal policy explain most of the observed difference in worker productivity across rich and poor countries?
13. Why do higher values for α and μ raise the long-run impact of tax rate differences across countries on worker productivity differences across countries? How can one justify higher values for α and μ than those used in the calibration exercise from the text?
14. What is the purpose of introducing the variable u? How might u vary across rich and poor countries?
15. Why does theory predict that growth due to *physical* capital accumulation naturally slows down? Does the same argument apply to human capital accumulation? Explain.
16. What has happened to the national saving rate across the developed countries since WWII? Offer an explanation that applies to all developed countries.
17. What is "crowding out?" Has crowding out occurred in developed countries such as the U.S.? Does crowding out apply to public as well as private capital?
18. Provide evidence that the growth rate of human capital has slowed.
19. Offer a reason we should be optimistic that technological progress will accelerate in the twenty-first century and a reason we should be pessimistic.
20. Use the results of this chapter to explain the growth facts *G1* and *G6*.

Problems

1. Solve the household maximization problem with fiscal policy and derive (3.6).
2. Derive the transition equation for private capital intensity under each type of government purchases policy, i.e. derive (3.9), (3.10), and (3.14).
3. Show the tax rate that maximizes the height of the transition equation given by (3.14) is $\tau^* = \mu(1 - \alpha)$. Hint: it simplifies things to take the natural log of the right-hand-side of (3.14) first. Taking the natural log is a monotonic transformation, so maximize the new expression is the same as maximizing the original expression.
4. Government Purchases—Consumption of Private Goods.

   For this question assume the following: A = 10, n = 1, d = 0, N = 100, $\beta = 1/2$, and $\alpha = 1/3$.

(a) Compute the initial steady state values of k, y, w, and Y with no government.
(b) Introduce the government but now assume $\varsigma = 0$ and $c_t^g = 2$. Starting from the initial steady state with no government, use (3.9) to compute the transition path for k_t, y_t, w_t, τ_t, and Y_t over the next 5 periods that results from introducing the government.
(c) Explain your answer to (b) using a transition equation figure.

5. Government Purchases—Employment.
   (a) Use the assumptions of *Problem* 4 to establish the same initial steady state without a government.
   (b) Now introduce the government with a work force equal to ςN, where $\varsigma = 0.10$, but assume no purchases of private consumption or investment goods. Starting from the initial steady state with no government, use (3.10) to compute the transition path for k_t, y_t, w_t, τ_t, and Y_t over the next 5 periods that results from introducing the government.
   (c) Explain your answer to (b).
6. Government Investment I
   Use the same assumptions as in Problem 4, but now let $\varsigma = 0$, $c_t^g = 0$, $\mu = 1/3$ and $\tau = 0.10$. Assume the initial capital-labor ratio is $k_0 = 0.05$. Using (3.14), compute the transition path for k_t, y_t, w_t, τ_t, and Y_t over the next 5 periods that results from introducing the government. Explain your answer.
7. Government Investment II
   Use the same assumptions as in Problem 4, but now let $\varsigma = 0$, $c_t^g = 0$, $\mu = 1/3$ and $\tau = 0.20$. Assume the initial capital-labor ratio is $k_0 = 0.05$. Using (3.14), compute the transition path for k_t, y_t, w_t, τ_t, and Y_t over the next 5 periods that results from introducing the government. Explain the difference between the transition paths in Problems 6 and 7.
8. Government Investment III
   Use the same assumptions as in Problem 4, but now let $\varsigma = 0$, $c_t^g = 0$, $\mu = 1/3$ and $\tau = 0.30$. Assume the initial capital-labor ratio is $k_0 = 0.05$. Using (3.14), compute the transition path for k_t, y_t, w_t, τ_t, and Y_t over the next 5 periods that results from introducing the government. Explain the difference in the three transition paths from *Problems* 6–8. *Hint*: Remember the lesson learned in *Problem* 3 and in the text.

   Base your answers to the next three questions on the model associated with the transition equation given by (3.14), where $y_t = Ag_t^{\mu(1-\alpha)}k_t^{\alpha}$. The next three Problems explore the tax rates that maximize steady state worker productivity and household utility, as discussed in the chapter Appendix.
9. Note that, for $E_t \equiv 1$, steady state worker productivity can be written as $\bar{y} = A\left[\frac{1+\beta}{\beta}\right]^{\mu(1-\alpha)}\left[\frac{\tau}{1-\tau}\right]^{\mu(1-\alpha)}\bar{k}^{\alpha+\mu(1-\alpha)}$. To derive the tax rate that maximizes steady state worker productivity complete the following steps.
   (i) Take the natural log of $\bar{y}$.

(ii) The expression in (i) involves the natural log of $\bar{k}$. You can write this expression in terms of the tax rate by solving for the steady state associated with the transition Eq. (3.27) and then taking the natural log.

(iii) Now the hard part. Collect terms that involve the tax rate and ignore other terms that will not be affected by the choice of the tax rate. This step is messy but you should end up concluding that maximizing worker productivity is equivalent to maximizing the expression, $\mu(1 - \alpha) \ln(\tau) + \alpha \ln(1 - \tau)$.

(iv) Maximize the expression from (iii) with respect to τ and solve for the tax rate.

10. Note that, for $E_t \equiv 1$, steady state utility can be written as

$$U = \ln\left((1-\tau)\bar{w}\bar{g}^{\mu}\right) + \beta \ln\left(\beta R(1-\tau)\bar{w}\bar{g}^{\mu}\right).$$

(i) Assume that $\delta = 1$, and write utility as

$$\begin{aligned} U &= (1+\beta)\ln(1-\tau) + (1+\beta)\ln\left(\bar{w}\bar{g}^{\mu}\right) + \beta\ln\beta + \beta\ln\bar{r} \\ &= (1+\beta)\ln(1-\tau) + (1+\beta)\ln((1-\alpha)\bar{y}) + \beta\ln\beta + \beta\ln\left(\frac{\alpha\bar{y}}{\bar{k}}\right). \end{aligned}$$

(ii) Use your analysis from *Problem* 29 to write out $\bar{y}$ and $\bar{k}$ in terms of the tax rate and other expressions. Collect all terms involving the tax rate and simplify. Very messy, but you should eventually conclude that maximizing utility is equivalent to maximizing the expression, $\mu(1 - \alpha)(1 + \beta) \ln(\tau) + (1 - \mu(1 - \alpha) + \alpha\beta) \ln(1 - \tau)$.

(iii) Maximize the expression from (ii) with respect to $\tau$and solve for the tax rate.

11. Make the following parameter assumptions: $\varsigma = 0$, $\beta = 1/2$, $\alpha = \mu = 1/3$. Compute the tax rates $\tau{*}{*}{*}$, $\tau{*}{*}$, and $\tau{*}$.

12. Note, from the relationships established in the text,

$$y_t = A g_t^{\mu(1-\alpha)} k_t^{\alpha} = A\left[\frac{\tau}{1-\tau}\frac{1+\beta}{\beta}\right]^{\mu(1-\alpha)} k_t^{\alpha+\mu(1-\alpha)}.$$

Also note, with no exogenous technological progress, the growth rate in y_t is

$$\frac{y_{t+1}}{y_t} - 1 = \left(\frac{k_{t+1}}{k_t}\right)^{\alpha+\mu(1-\alpha)} - 1.$$

To get the *annualized* growth rate under the interpretation that each period lasts 30 years, write

$$\left(\frac{y_{t+1}}{y_t}\right)^{1/30} - 1 = \left(\frac{k_{t+1}}{k_t}\right)^{(\alpha+\mu(1-\alpha))/30} - 1$$

The growth rate expressions applies equally well to an economy without the government, as was the case in Chap. 2, by setting $\mu = 0$.

Use the growth rate formula to compute the annualized growth rates associated with *Problem* 13 of Chap. 2 (where $\mu = 0$). Next, use the growth rate formula to compute the annualized growth rates associated with *Problem* 7 of this chapter (where $\mu = 1/3$). Compare the growth rates and explain the difference. What would happen if μ were greater than 1/3?

13. Suppose the governments in two locations (countries, cities, regions) provide the same value of g_t. The two locations, A and B, are otherwise identical except the population size in location B is twice that of location A. If $\xi = \mu = 1/3$, what is the ratio of TFP in location B relative to location A?
14. In a small economy that is perfectly open to private capital flows, with $\alpha = \mu = 1/2$, determine as accurately as you can what happens to k_t and y_t if
    (a) A doubles
    (b) g doubles
    (c) r* doubles.
15. Solve the household maximization problem from Sect. 3.6 to get (3.22).
16. Use (3.20, 3.21 and 3.22) to derive the transition Eq. (3.24).
17. Derive the steady state expressions (3.29a, 3.29b, and 3.29c).
18. Prove that τin (3.27a) is decreasing in ϕ.
19. In the following scenarios, consider the impact of fiscal policy differences on worker productivity differences across countries, using the model from Sect. 3.6. In each scenario assume that $\tau^R = 0.15$, $\tau^P = 0.35$, and $\alpha = 1/3$.
    (a) Assuming that $u = 0$, what value must μ take for fiscal policy to explain a two-fold difference in steady state worker productivity across rich and poor countries?
    (b) If $u = 0.5$ and $\mu = 1/2$, what is the difference in worker productivity across countries?

Appendix

Tax Rates

In the text, we consider the value of the wage tax rate that maximizes the height of the transition equation for the private capital-labor ratio. Maximizing the growth in private capital intensity is not necessarily a reasonable objective. Instead we might consider the tax rate that maximizes state worker productivity ($\tau{*}{*}{*}$) or steady state household utility ($\tau{*}{*}$). One can compute these tax rates as well (see *Problems* 9–11). The comparison of the three tax rates is

$$\tau{*}{*}{*} = \frac{\mu(1-\alpha)}{\mu(1-\alpha)+\alpha} >$$

$$\tau** = \frac{\mu(1-\alpha)}{\mu(1-\alpha)+\alpha}\left(\frac{1+\beta}{\left(\frac{1}{\mu(1-\alpha)+\alpha}\right)+\beta}\right) = \mu(1-\alpha)\left(\frac{1+\beta}{1+\beta(\mu(1-\alpha)+\alpha)}\right) >$$

$$\tau* = \mu(1-\alpha),$$

because $\mu(1-\alpha)+\alpha < 1$.

The tax rate that maximizes steady utility is perhaps the most compelling. It is higher than the tax rate that maximizes steady state capital intensity because there is a benefit to households of keeping the private capital intensity lower than the maximum. All households are savers, so a higher return to capital, other things constant, raises household welfare. The desire to keep the return to capital high creates an incentive to keep private capital intensity low. This consideration causes the policy maker to set the tax rate higher than the one that maximizes the steady state value of k.

The highest tax rate is the one that maximizes steady state worker productivity. This tax rate is higher than the rate that maximizes steady state utility because it does not account for the fact that a higher tax rate on wages lowers the after-tax wage that determines household consumption and instead only focuses on the before-tax wage associated with worker productivity.

References

Abel, J., and Deitz, R., 2014, “Do the Benefits of College stiil Outweigh the Costs?” *Current Issues*, New York: Federal Reserve of New York.

Adams, C., 2016, “2015, SAT, ACT Suggest that Many Students are Not College-Ready,” *Education Week*, 35(3),6.

Aghion, P., Belsey, T., Browne, J., Caselli, F., Lambert, R., Lomax, R., Pissarides, C., Stern, N., and Van Reenen, J., 2013, *Investing for Prosperity*, Report of the LSE Growth Commission.

Alfaro, L., 2016, “Gains from Foreign Direct Investment: Macro and Micro Approaches,” *World Bank Economic Review*: *Advanced Access*, March 23, 2016.

Ashauer, D., 1989, “Is Public Expenditure Productive?” *Journal of Monetary Economics*, 23, 177–200.

Ashauer, D., 2000, “Public Capital and Economic Growth: Issues of Quantity, Finance, and Efficiency,” *Economic Development and Cultural Change*, 48, 391–406.

Bennett, W., and Wilezol, D., 2013, *Is College Worth It*, Nashville: Thomas Nelson.

Bivens, J., 2012, “Public Investment: The Next New Thing for Powering Economic Growth,” *EPI Briefing Paper #338*, Washington Economic Policy Institute.

Brynjolfsson, E., and McAfee, A., 2014, *The Second Machine Age*, New York: Norton.

Carneiro, P., and Heckman, J., 2005, “Human Capital Policy,” in J. Heckman and A. Krueger, *Inequality in America*, Cambridge Mass.: MIT Press.

Chakraborty, S. and Dabla-Norris, E., 2011, “The Quality of Public Investment,” *B.E. Journal of Macroeconomics*, 11, 1–29.

Collier, P., 2009, *War, Guns, and Votes*, New York: Harper Collins.

Dobrescu, Loretti I. & Kotlikoff, Laurence J. & Motta, Alberto, 2012. “Why aren’t Developed Countries Saving?,” *European Economic Review*, 56(6), 1261–1275.

Friedman, T., and Mandlebaum, M., 2012, *That Used to be Us*, Picodar: New York.

Gloom, G. and Ravikumar, B., 1997, "Productive Government Expenditures and Long-run Growth," *Journal of Economic Dynamics and Control*, 21, 183–204.

Gordon, R., 2016, *The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War*, Princeton University Press: Princeton.

Hanushek, E., 2005, "Comments," in J. Heckman and A. Krueger, *Inequality in America*, Cambridge Mass.: MIT Press.

Heckman, J., and Krueger, A., 2005, *Inequality in America: What Role for Human Capital Policies?*, Cambridge, Mass.: MIT Press.

Heckman, J., Moon, S., Pinto, R., Savelyev, P., and Yavitz, 2010, The Rate of Return to the HighScope Perry Preschool Program, *Journal of Public Economics*, 94, 114–129.

Hunt, J., and Gauthier-Loiselle, M., 2010, "How Much Does Immigration Boost Innovation," *American Economic Journal: Macroeconomics*, 2, 31–56.

Ivanyna, M., Mourmouras, A., and Rangazas, P., 2018, *The Macroecoomics of Corruption*, Switzerland: Springer.

Jones, C., 2002, "Sources of U.S. Economic Growthin a World of Ideas," *American Economic Review*, 92(1), 220–239.

Jones, C., and Vollrath, D., 2013, *Introduction to Economic Growth*, New York: W.W. Norton.

Katz, L., 2005, "Comments," in J. Heckman and A. Krueger, *Inequality in America*, Cambridge Mass.: MIT Press.

Kotlikoff, L., 2015, "America's Fiscal Insolvency and its Generational Consequences," Testimony to the Senate Budget Committee, February 25, 2015.

Kose, M., Prasad, E., Rogoff, K., and Wei, S., 2009, "Financial Globalization: A Reappraisal," *IMF Staff Papers*, 56 (1), 8–62.

Kraay, A., and McKenzie, D., 2014, "Do Poverty Traps Exist? Assessing the Evidence," *Journal of Economic Perpsectives*, 28(3), 127–0145.

Ljungquist, L., and Sargent, T., 2004, *Recursive Macroeconomic Theory*, Cambridge Mass.: MIT Press.

Malinovskaya, A., and Wessel, D., 2017, "The Hutchins Center Explains: Public Investment," *Upfront*, Washington DC: Brookings Institution.

Manuelli, R., and Seshadri, A., 2014, "Human Capital and the Wealth of Nations," *American Economic Review*, 104 (9), 2736–2762.

Mulligan, C., and Tsui, K., 2015, "Political Entry, Public Policies, and the Economy," *Research in Economics*, 69, 377–397.

National Center for Education Statistics, 2015, "Immediate Enrollement Rates," *The Condition of Education 2015*, Washington: U.S. Department of Education.

OECD, 2014, *Policy Challenges for the Next Fifty Years*, Economic Policy Paper No. 9, Paris; OECD Publishing.

OECD, 2015, *The Future of Productivity*, Paris: OECD Publishing.

Parente, S. and Prescott, E., 2000, *Barriers to Riches*, Cambridge Mass.: MIT Press.

Putnam, R., 2015, *Our Kids*, New York: Simon and Schuster.

Rangazas, P. , 2002, The quantity and quality of schooling and U.S. labor productivity growth (1870-2000). *Review of Economic Dynamics* 5(4): 932–964.

Rothwell, J., 2016, "The Declining Productivity of Education," *Social Mobility Memo*, Washington: Brookings Institution.

Stupak, J., 2018, "Economic Impact of Infrastructure Investment," *Congressional Research Service*, www.crs.gov.

Turner, S. 2004, "Going to College and Finishing College: Explaining Different Educational Outcomes" in Caroline Hoxby editor *College Decisions: How Students Actually Make Them and How They Could*, Chicago : University of Chicago Press.

Viig, J., 2011, *The American Technological Challenge*, Algora Press: New York.

Wadhwa, V., 2012, *The Immigrant Exodus*, Philadelphia: Warton Digital Press.

Schooling and Fertility

4

This chapter is motivated by two of the growth facts stated in Chap. 1.

G2—Children spend more time in school, within and across years, and less time working as an economy develops

G3—Population growth rates may first rise but eventually experience a steady decline as economies develop

We have seen evidence for *G2* from Table 2.2 in Chap. 2. The school-time investment is depicted in Fig. 4.1. In the United States, time spent in school by children aged 0–19 years expanded more than 3-fold from 1870 to 1990. There is some indication that time spent in school was less earlier in the nineteenth century than in 1870 (Lord and Rangazas (2006, Table 2)), suggesting a larger increase over the two centuries that the United States experienced modern growth.

The decrease in population growth stated in *G3* is known as the *demographic transition*. The demographic transition is a byproduct of a race between expanding longevity and declining fertility. During a country's economic development, the decline in fertility eventually wins the race, pulling down population growth. Over the nineteenth and twentieth centuries, the average number of children per woman in the United States fell dramatically from 7 to 2 (Haines (2000)).

Why does fertility decline over the course of development and what does it have to do with the rise in schooling? We extend the model from Chap. 2 to explain both facts. The extended model is based on the theory of schooling and fertility created by the Nobel Prize winning economist, Gary Becker. Becker argues that parents enjoy having children and helping them become productive adults. However, this creates a tradeoff between the *quantity* (number) and *quality* (adult productivity) of children because the more parents invest in children the more costly it is to raise a child. One dimension of child quality is how much schooling they receive. Figures 4.1 and 4.2 reveal that the quantity and quality of children tend to be inversely related; as schooling rises, fertility falls. Becker argues that the inverse relationship is due to the rising costs of children associated with the increase in schooling.

S. Das et al., *Economic Growth and Development*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-319-89755-4_4

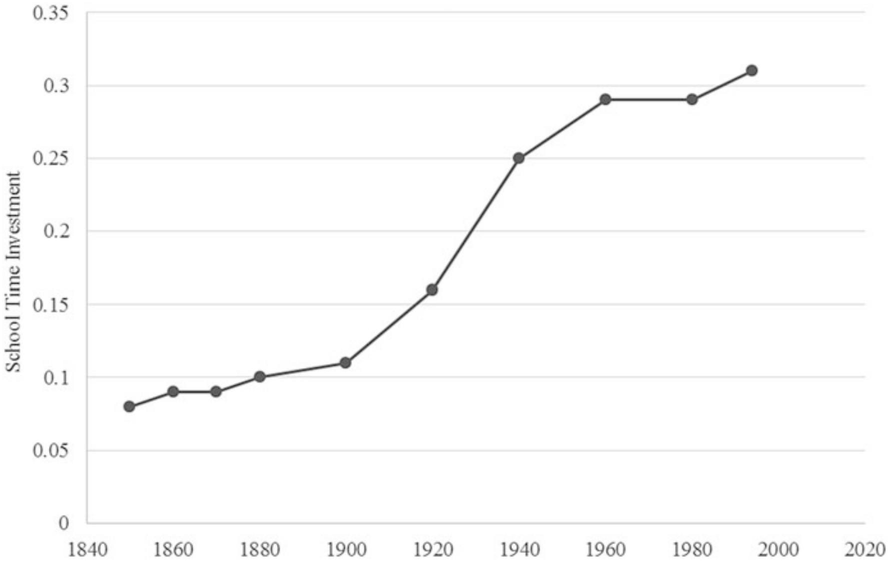


Fig. 4.1 The rise in time spent in school

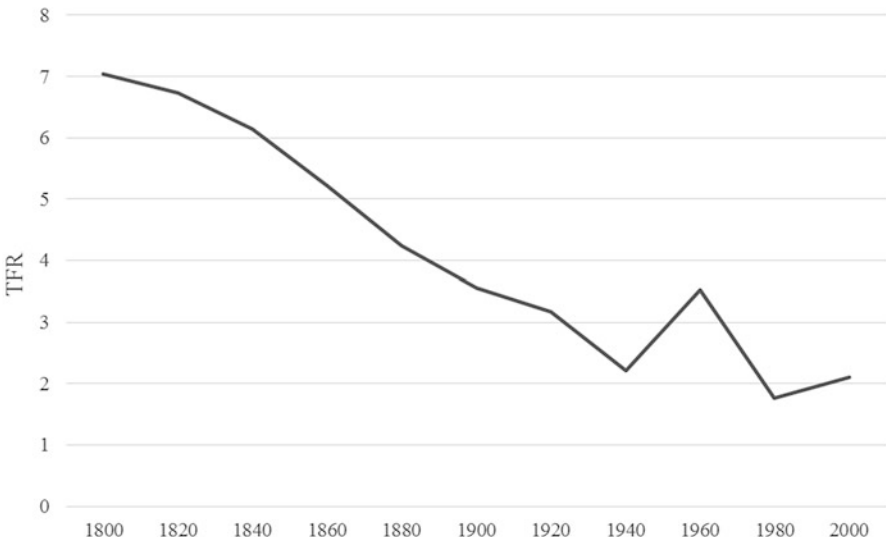


Fig. 4.2 The fall in fertility

The extended model introduces two new endogenous determinants of economic growth. Schooling is now an endogenous variable, explaining the rise in human capital from Chap. 2 that was shown to be important in explaining growth in the United States history. Chapter 2 also identified that population growth lowers economic growth because it dilutes physical capital intensity. Population growth is now linked to other economic variables through the theory of fertility. The extended model also introduces the concept of a *poverty trap*—a low income steady state, with high fertility and low schooling, that prevents full-fledged modern growth from taking off.

4.1 The Quantity and Quality of Children

This section develops a theory of fertility and schooling. In addition to choosing consumption over their lifetime, households also choose the number of children in the family and how much schooling to invest in each child. In Chap. 2, household consumption and saving decisions provide the foundation for growth via physical capital accumulation. The household choices of fertility and schooling not only allow a unified explanation for *G2* and *G3*, but also give a theory of human capital accumulation and an extended theory of physical capital accumulation that includes population growth.

4.1.1 Households

Households are viewed as living for three periods. The three periods correspond to one period of childhood and two periods of adulthood. This is simply the two-period life-cycle model with childhood now explicitly recognized as a third period. The reason for focusing on childhood is that, in addition to providing a theory of saving, we now introduce endogenous fertility and schooling decisions made by parents.

As in Chap. 2, households value their consumption over the two periods of adulthood (c_{1t}, c_{2t+1}). They also value the *quantity* (n_{t+1}) and *quality* of their children, as measured by the child's adult earnings ($w_{t+1}D_{t+1}h_{t+1}$). Valuing a child's adult earnings is similar to the "warm glow" preference for intergenerational transfers from Chap. 2. The child's adult earnings are the product of the after-tax market rental rate for skills (w_{t+1}), the productivity index (D_{t+1}), and embodied skills, or human capital (h_{t+1}) of the worker.

Formally, the preferences of parents are given by

$$U_t = \ln c_{1t} + \beta \ln c_{2t+1} + \psi \ln \left(n_{t+1} w_{t+1} D_{t+1} h_{t+1}\right) \tag{4.1}$$

where $0 < \beta < 1$ and $\psi > 0$ are preference parameters.[1] The new parameter ψ gauges the value placed on children relative to family consumption. Parents directly choose n_{t+1}. They affect their child's adult productivity and earnings by

[1] Galor and Moav (2002) generalize this specification by allowing for a separate utility weight on the quantity and quality of children. They then go on to develop an evolutionary theory in which households raise the weight they placed on the quality of their children over the course of economic development. Using this more flexible specification would increase the ability of our model to fit the stylized growth facts.

choosing the time the child spends in school, e_t. The adult human capital of the child is given by $h_{t+1} = e_t^{\theta}$, where $0 < \theta < 1$ is a parameter that captures the effect of schooling on human capital accumulation.

For simplicity we ignore purchased goods and services inputs used in schooling (e.g. as reflected in tuition and fees) and focus only on student time because we want to study the schooling versus child labor tradeoff featured in growth fact *G2*. *Problems* 2–4 introduce goods inputs and explore their implications for those interested.

As in Chap. 2, adults inelastically supply one unit of labor when young and zero units when old. Children have an endowment of $T < 1$ units of time that they can use to attend school (e_t) or work ($T - e_t$). Children have less than one unit of time to spend productively because early in childhood they are too young to either attend school or to work, and in the middle years of childhood they do not have the mental or physical endurance to learn or work as long as an adult.

We think of the children as being too young to work over the early part of their lives or that a minimum amount of schooling is needed for the child to be productive. Under either interpretation, each child invests at least $\bar{e}$ units of time into learning during the first portion of their childhood. This gives older children $\gamma\bar{h}_t = \gamma\bar{e}^{\theta}$ units of human capital that can be used in production during the later years of childhood, where $0 < \gamma < 1$ reflects the fact that even older children lack the relative physical strength or experience in applying knowledge to production compared to an adult. Thus, per hour of work, a person is more productive in adulthood than in childhood because of greater strength and experience ($1 > \gamma$) and possibly additional schooling received later in childhood ($e_t \geq \bar{e}$).

While children may work as they become older, providing income to the family, they are also expensive to care for and feed. To raise each child requires a loss of adult consumption equal to a fixed fraction η of the adult's first period wages. One can interpret the cost of raising a child as (i) the parent's forgone wages associated with the time away from work needed watch over and informally educate a young child or (ii) the loss in adult consumption associated with providing consumption goods to children. In this chapter, either interpretation is fine. In future chapters we may find it convenient to stress one interpretation over the other.

If we put all these elements together we have the following interrelated cost concepts associated with the decision to (i) have a child and (ii) send the child to school:

(i) *net cost of raising a child, forgone adult consumption minus child income*

$$\eta w_t D_t h_t - w_t D_t \gamma \bar{h}(T - e_t)$$

(ii) *cost of time spent in school, forgone child wages*

$$w_t D_t \gamma \bar{h}$$

The new choices of the quantity and quality of children, and the associated costs, are included in the family's lifetime budget constraint,

$$c_{1t} + \frac{c_{2t+1}}{R_t} + n_{t+1}\eta w_t D_t h_t = w_t D_t h_t + n_{t+1} w_t D_t \gamma \bar{h}(T - e_t). \tag{4.2}$$

The new terms introduced by this chapter's extensions are the forgone adult consumption associated with raising children ($n_{t+1}\eta w_t D_t h_t$) and the family income generated by child labor ($n_{t+1} w_t D_t \gamma \bar{h}(T - e_t)$). The net cost of raising children is the difference in these two expressions.

Parents choose consumption, the number of children, and the schooling/work of each child to maximize (4.1) subject to (4.2). An important detail of parents' decision making is that children must at least spend $\bar{e}$ units of time in learning during the first portion of their childhood. The economic fundamentals may make parents "wish" their children could go to school less than $\bar{e}$ units of time and work more, but this is not possible. Schooling time can never fall below $\bar{e}$ and this constraint must be accounted for.

The demand functions for children, schooling, and assets used to finance retirement consumption that result from maximizing utility are, see *Problem* 1,

$$\begin{aligned} n_{t+1} &= \frac{\psi w_t D_t h_t}{(1+\beta+\psi)(\eta w_t D_t h_t - w_t D_t (T - e_t)\gamma \bar{h})} \\ &= \frac{\psi}{(1+\beta+\psi)\left(\eta - \gamma(T - e_t)\left(\bar{e}/e_{t-1}\right)^{\theta}\right)} \end{aligned} \tag{4.3a}$$

$$\begin{aligned} e_t &= \max\left[\frac{\theta}{(1-\theta)} \frac{\eta w_t D_t h_t - w_t D_t T \gamma \bar{h}}{w_t D_t \gamma \bar{h}}, \bar{e}\right] \\ &= \max\left[\frac{\theta\left(\eta\left(e_{t-1}/\bar{e}\right)^{\theta} - \gamma T\right)}{\gamma(1-\theta)}, \bar{e}\right] \end{aligned} \tag{4.3b}$$

$$s_t = \left[\frac{\beta}{1+\beta+\psi}\right] w_t D_t h_t. \tag{4.3c}$$

From (4.3a), we see that fertility is positively related to adult income (numerator)—an income effect and negatively related to the net cost of children (denominator)—a "price" effect. After some algebraic simplification, fertility is shown to be affected by three important variables.

4.1.2 Determinants of Fertility

(i) *Relative productivity of children* (γ)—the greater the relative productivity of children, the lower is their net costs and the greater is fertility

(ii) *Parents' schooling* (e_{t-1})—the greater is the schooling of parents the higher the opportunity cost of raising a child, reducing fertility
(iii) *Child's schooling* (e_t)—the greater the schooling of children the less work they do, increasing the net cost of children, reducing fertility

Equation (4.3b) says schooling is positively related to the minimum net cost of children (the net cost when children work as soon as they can, found in the numerator) and negatively related to the forgone child earnings associated with schooling (denominator). Schooling is high when children are expensive to raise and forgone child earnings are low. However, if children are sufficiently cheap and forgone child earnings are high, then fertility is high and parents want only the minimum schooling for each child. The economic fundamentals must be such that parents want schooling to exceed $\bar{e}$. It is only in this case that we can use the equation, $e_t = \theta\left(\eta\left(e_{t-1}/\bar{e}\right)^{\theta} - \gamma T\right)/\gamma(1-\theta)$.

4.1.3 Determinants of Schooling ($e_t > \bar{e}$)

(i) *Relative productivity of children* (γ)—the greater the relative productivity of children, the larger are the forgone earnings from sending a child to school and the less schooling each child receives
(ii) *Parents' schooling* (e_{t-1})—the greater is the schooling of parents the higher the opportunity cost of raising a child, increasing schooling as parents substitute quality for quantity

Equation (4.3c) gives a saving function of the same form that was derived in Chap. 2. The presence of schooling and fertility choices does not alter the saving behavior of the household, except that the precise fraction of adult wages that is saved is now affected by the taste for children parameter, ψ. The higher the value of ψ, the more resources are spent on children and the smaller the fraction of wages that is saved.

4.2 The Nature of the Fertility-Schooling Interaction

This section provides a detailed discussion of some of the properties and implications of the schooling-fertility theory from Sect. 4.1. We in particular stress the nature of the interaction between fertility and schooling that underlies an explanation for the growth facts *G1* and *G2*.

4.2.1 Schooling and Fertility Are Independent of Other Variables

From (4.3a) and (4.3b), we see that the schooling and fertility choices are independent of technological progress (D) and the rental rates on physical (r) and human capital (w). It is not too surprising that the rental rate on physical capital does not affect the first period choices of the quantity and quality of children. However, it is surprising that higher wages (wD) have no effect. A higher productivity index and higher rental rate paid to human capital does raise the earnings of both parents *and* children but does *not* affect the *relative* earnings or the net cost of children. This result is important in three ways.

First, the sources of growth stressed in Chap. 2 do not affect the new choices of this section. The fertility-schooling interaction that explains *G1* and *G2* is determined independently from the dynamics of the standard neoclassical growth model—an interesting result and a convenient analytical simplification. A comprehensive attempt to explain growth can decompose the sources of growth by starting with fertility and schooling, and then move to physical capital accumulation. Fertility and schooling *do* affect physical capital accumulation, even though the reverse is not true, as is made explicit in Chap. 5 when the complete one-sector model is presented.

Second, schooling can be determined in isolation. Equation (4.3b) is a transition equation for schooling alone. Given an initial value for parent's schooling, an entire path of schooling into the future is determined. Given the path of schooling, the path of fertility can then be determined. This means explaining growth has the following *recursive* structure.

Schooling => Fertility => Physical Capital

Finally, the fact that technological progress and physical capital accumulation do not affect schooling increases the chance for a *schooling trap* that limits a country's growth. As will be discussed further below, the fundamentals of the economy may cause parents to prefer the minimum schooling for their children, $e_t = \bar{e}$. Growth in the economy for other reasons will not push the economy out of this schooling trap. The trap must be addressed more directly.

4.2.2 Changes in Fertility Across Time and Households

From (4.3a) we see that for fertility to fall, there must be a rise in schooling. An increase in schooling raises the wages of parents relative to their children, which raises the net cost of children and lowers fertility. This is crux of our explanation for *G3*. Whatever the forces that initially push fertility and population growth rates up as economies develop, we will explore some of these below and in Chap. 6, eventually the rise in schooling dominates and pulls fertility and population growth down. As demonstrated in Chap. 2, the fall in population growth is important in allowing physical capital accumulation per worker to intensify and speed growth in per capita incomes.

It should be recognized that the key to the fertility decline is an increase in the schooling of *older* children. An increase in schooling of older children raises the net cost of children and lowers fertility, creating the "quantity-quality" tradeoff made famous by the Nobel-Prize winning economist, Gary Becker (see Becker (1960, 1981) and Becker and Lewis (1973)). Less well known is that Becker acknowledged that the interaction between schooling and fertility could be *positive*.

> The net cost of children is reduced if they contribute to family income by performing household chores, working in family business, or working in the market place. Then an increase in earning potential of children would increase the demand for children (Becker (1981, p. 96)).

Increased schooling of *young* children, a rise in $\bar{e}$ due to an increase in days attending school of young children who do not yet have the capacity to work, clearly increases their earning potential when they are older and able to generate income for the family. This type of rise in schooling *lowers* the net cost of children and increases fertility. Expanded schooling for younger children is one reason that fertility may initially rise over the course of development.

It is also important to notice that *cross-sectional variation* in schooling and fertility across different households in a given period can work very differently than changes in *average* schooling and fertility for an entire country *over time*. Equations (4.3a) and (4.3b) indicate that while ψ affects fertility it does not affect schooling. There is naturally *variation in* ψ across households that causes fertility to vary across households with no change in schooling per child. In Chaps. 7 and 8 we find that the same result holds for *variation in non-labor income* across households. Higher non-labor income leads to greater fertility with no change in schooling per child. This is consistent with empirical findings using cross-sectional data sets that children raised in smaller families receive no more schooling (e.g. Banerjee and Duflo (2011, p. 108)). These findings do not reject the quantity-quality tradeoff generated by a causal link from the schooling of older children to fertility, for given values of ψ or for a given distribution of ψ across all households in an entire economy. The inverse relationship between the quantity and quality of children is clearly present in the time series evidence for both historical and currently developing countries (e.g. Lord and Rangazas (2006) and Fernihough (2017)).

4.2.3 The Relative Productivity of Children

A key parameter affecting both schooling and fertility is the relative productivity of working children (γ). A rise in γ decreases schooling and raises fertility because it raises the opportunity cost of schooling and lowers the net cost of children. Variation in γ in cross-sectional or time series data can explain variation in both schooling and fertility.

It is clear that the value of γ may change across regions or countries. For example, children are more productive in the "lighter" farming of crops grown in warmer climates (e.g. rice, cotton, sugar) than in the "heavy" farming of colder climates (e.g. wheat, corn, barley). Thus, one would expect less schooling and greater fertility in warmer climates.

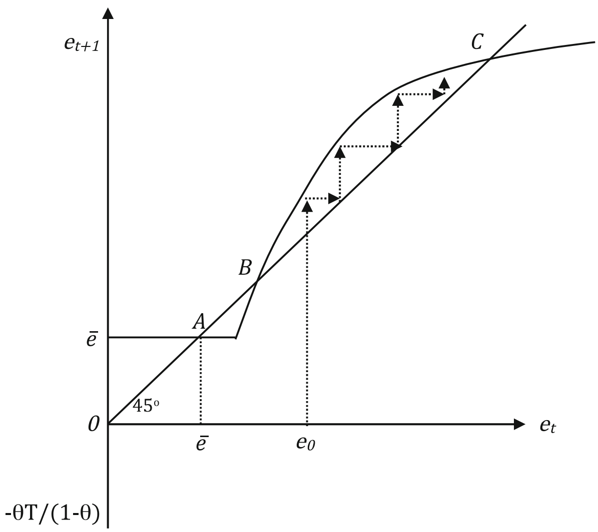
Technological change can be associated with increases or decreases in γ over time. For example, the early Industrial Revolution involved the introduction of new equipment for making textiles that could be operated by children without work experience, causing a rise in γ. This could explain why fertility initially rose in the early stages of historical modern growth.

4.3 The Schooling Poverty Trap and Schooling Dynamics

Equation (4.3b) is a transition equation similar to those we encountered for k_t in Chaps. 2 and 3. The equation relates e_t, current schooling of children, to e_{t-1}, past schooling that the current generations of parents received when they were children. A key difference between the schooling transition equation and the transition equation for k_t is the importance of initial conditions. For schooling to begin to rise, the initial generation of parents must have human capital that is sufficiently higher than the human capital of their working-age children. If this condition holds, schooling will rise each period and growth in human capital ensues. As schooling rises, the effect of a given increment in schooling has a diminishing effect on human capital formation and wages. With the right initial conditions, the transition equation given by (4.3b) will eventually exhibit the standard properties of neoclassical growth as human capital accumulates.

We can depict the dynamic nature of the schooling equation in the diagram displayed in Fig. 4.3. The horizontal axis keeps track of the parent's schooling and the vertical axis keeps track of the child's schooling. The curve is a plot of (4.3b), the schooling equation that relates schooling across generations. The lowest value that

Fig. 4.3 Schooling dynamics



e_{t+1} can take is $\bar{e}$ rather than at the origin as it was with the physical capital transition equation from Chap. 2. The graph intersects the 45° line three times, at points *A*, *B* and *C*, but only *A* and *C* are dynamically stable. Starting away from *B* by the slightest amount will send the economy to either point *A* or point *C*.

The unusual configuration of the transition equation is created by the existence of $\bar{e}$, the minimum level of schooling for young children, drawn here to be less than the schooling level at point *B*. A point like *B* means parents with education greater than $\bar{e}$ may choose $\bar{e}$ for their children, a possibility that cannot be ruled out in theory (regress in education over time). For schooling to increase over time, the schooling level for parents must be to the right of *B* or, more generally, greater than the maximum of the schooling levels associated with point *B* and $\bar{e}$.

Starting to the right of *B* will cause schooling to rise, but in relatively small increments initially. As schooling rises, the increments in schooling across generations become larger until the economy nears the stable steady state at *C*, where the increments once again become smaller and converge to zero. So, *provided that schooling is sufficiently high initially*, the model predicts relatively small increments in schooling initially, an acceleration of schooling in the middle of the transition, and then a slowdown as the steady state is approached. However, there is no guarantee that the economy has the proper initial conditions to generate any growth in human capital. If the economy's initial human capital investments are less than those associated with point *B*, it will eventually be stuck in a *poverty trap* where schooling remains at $\bar{e}$ indefinitely, the other stable steady state located at point *A*.

A couple of additional points are worth noting about the poverty trap. First, the rate of return to schooling may be quite high at $\bar{e}$. Parents nevertheless decide that they cannot afford to forgo the family income that would be lost if older children worked less and spent more time in school. If the rate of return to schooling exceeds the market interest rate, then this situation is inefficient because in principle the family could borrow to cover the forgone earnings and then collect more than enough from the enhanced adult earnings of their children to pay back the loan with interest. However, as discussed in Sect. 2.6 of Chap. 2, such an outcome requires that well-functioning *intergenerational* loan markets exist, where parents assume a debt on behalf of their dependent children and the lender is able to collect the loan repayment from those same children when there are adults. There are so many difficult incentive and legal issues associated with this type of transaction that intergenerational loan markets do not even exist in developed, let alone developing, countries. Thus, when the rates of returns to schooling are high at $\bar{e}$, some type of policy action that raises schooling is justified in order to improve productive efficiency.

Second, similar poverty traps may exist for physical capital. If there is a significant fixed cost associated with establishing a firm, then borrowing in credit markets becomes important for physical capital accumulation as well. It is generally believed that borrowing to finance physical capital is an easier financial transaction than for human capital because it does not require an intergenerational loan market and because physical capital may be confiscated if the loan is not repaid. However, at early stages of development these types of loans may also be quite costly. As a separate matter, the economic environment may be such that physical capital simply cannot compete with land as a productive asset. In Chap. 6, we examine a model where the state of technology must be sufficiently high before production based on physical capital is profitable (even when credit market for physical capital loans exist). In this setting, countries with governments and interest groups that block technological advancement can get stuck in poverty traps where all production takes place using labor and land.

4.4 Numerical Example

To reinforce the key features of the model, let's work through an extended numerical example. Start by assuming

(i) $\theta = 0.50$,

so that the human capital production function is

$$h_t = e_{t-1}^{1/2} = \sqrt{e_{t-1}}. \tag{4.4}$$

This assumption implies the transition equation for schooling can be written as,

$$e_t = \frac{\eta}{\gamma}\sqrt{\frac{e_{t-1}}{\bar{e}}} - T \quad \text{provided}\, e_t > \bar{e}.$$

To make things even more transparent, let's simplify further and assume that

(ii) $T = 0.50$
(iii) $\bar{e} = 1/9$.

The schooling equation then becomes

$$e_t = 3\frac{\eta}{\gamma}\sqrt{e_{t-1}} - 0.5, \quad \text{provided}\, e_t > \bar{e}. \tag{4.5}$$

With this transition equation, if you are given an initial schooling level, and values for the remaining parameters η and γ, you can numerically trace the entire path of schooling into the future. There are *Problems* at the end of the chapter that will ask you to do just that.

We can also simplify the plot of the transition equation given in Fig. 4.3 by assuming the special case where (a) the concave portion of the transition equation crosses the 45-degree line at $\bar{e}$ and (b) the concave portion of the transition equation does not cross the 45-degree line again until *after* the maximum schooling level of T. This gives us the configuration in Fig. 4.4. It has two flat portions representing the minimum level of schooling, $\bar{e}$, and the maximum level of schooling, T.

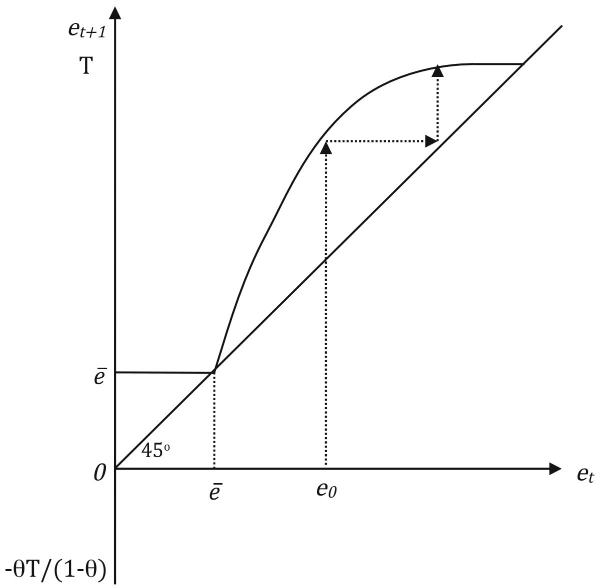


Fig. 4.4 Simplified schooling transition equation

In Fig. 4.4, if the parents start with the minimum schooling, their child also receives the minimum schooling—the schooling poverty trap. Every generation is stuck with the minimum level of schooling. To escape the poverty trap, parents must have more than the minimum level of schooling, i.e. e_{t-1} must be greater than $\bar{e}$ for some generation of parents. If this happens, schooling will take-off and grow over time until children are attending school for the full time endowment.

To trace the escape from the poverty trap, follow the arrows. In period 1, assume parents have a particular value of e_{t-1}, say e_0. Follow the arrow up to the curve representing the transition equation to find out what schooling the child receives on the vertical axis. Label this value e_1. Next, follow the horizontal arrow back to the 45-degree line and look down to the horizontal axis—this determines the schooling that parents possess in the next period, also e_1—the education of the child in period 1 *is* the education of the parent in period 2. In period 2, if parents have schooling time e_1, their children receive schooling time e_2. Proceeding in this way determines the entire dynamic path until the economy reaches the maximum schooling level (a second type of steady state).

The situation depicted in Fig. 4.3 is worse for growth than in Fig. 4.4. You can think of Fig. 4.3 as being created by an increase in γ that shifts the transition equation in Fig. 4.4 downward. The downward shift implies the schooling equation intersects the 45-degree line at a higher value of e_{t-1}, as depicted at point B of Fig. 4.3. The new steady state at point B, say $\widehat{e}$, is greater than $\bar{e}$. Why is this new steady state bad? It means that parents must now have schooling levels significantly greater than $\bar{e}$ to escape the poverty trap. If the parent's schooling is anywhere less than $\widehat{e}$, the level of schooling will *decline* over generations until the economy reaches $\bar{e}$. The declining school path is determined by the same procedure used to depict the rising school path in Figs. 4.3 and 4.4. To get on an increasing school path, parent's schooling must be greater than $\widehat{e}$, a value higher than $\bar{e}$.

4.4.1 Tracking Fertility

The fertility equation can also be simplified using the same assumptions made for the schooling equation,

$$n_{t+1} = \frac{\psi}{\eta - \frac{\gamma}{3}\frac{0.5 - e_t}{\sqrt{e_{t-1}}}}. \tag{4.6}$$

If schooling takes off and starts growing, fertility falls as greater schooling for parents raise the forgone wages and consumption associated with raising children and greater schooling of children reduces the income they bring to the family, both causing the cost of children to rise over time. If schooling is stuck in the poverty trap, so is fertility. The number of children in the fertility trap, when $e_t = \bar{e} = 1/9$, is

$$n_{t+1} = \frac{\psi}{\eta - 0.3899\gamma} \tag{4.7}$$

In contrast, if the country escapes the schooling trap, and schooling eventually reaches the maximum of T, fertility is

$$n_{t+1} = \frac{\psi}{\eta} \tag{4.8}$$

Let's continue to play with some numbers. Note that, under our assumptions, in the poverty trap (4.5) becomes $1/9 = 3\frac{\eta}{\gamma}\sqrt{1/9} - 0.5$, implying that $\frac{\eta}{\gamma} = 0.6111$. Empirical evidence suggests that a reasonable value for γ is 0.30 (see Lord and Rangazas (2006)). Using this estimate, we have $\eta = 0.30 \times 0.6111 = 0.1833$.

If we take the ratio of the maximum fertility, in the poverty trap, to the minimum fertility, in the equilibrium with maximum schooling, we get

$$\frac{\eta}{\eta - \gamma(T - \bar{e})} = \frac{0.1833}{0.1833 - 0.30 \times 0.39} = 2.75.$$

This means that fertility per household for poor countries in the schooling-fertility trap is almost 3 times higher than fertility per household in a developed country with high levels of schooling and no child labor. This is a realistic outcome as fertility in poor countries is often 6 children per woman compared to 2 children per woman in rich countries. *Problems* 7 and 8 continue the numerical analysis of fertility.

4.5 Schooling Poverty Traps: A Closer Look

This brings us to the question of why some countries escape the poverty trap and others do not. One reason is *culture*. Culture, for our purposes, is a collection of non-economic features of society that standard economic models do not directly capture. Culture can certainly have an effect on education. Basic literacy advanced in some countries, but not others, well before the onset of sustained modern growth. Educating the general population could be motivated by social cohesion, enlightenment, religion, or military goals, as well as by economic considerations. Cultural and political differences in the pursuit of these broader goals caused the timing and extent of education to vary across countries before modern economic growth began. In some countries literacy was high at the onset of growth, Germany and Sweden, and in other countries it was low, England.

In our model, we can think of Germany and Sweden starting to the right of the low schooling steady state. The subsequent growth in schooling helped accelerate their economic growth. England, on the other hand, was stuck in a low schooling poverty trap in the early stages of its growth. It needed child labor laws and compulsory schooling legislation during the early nineteenth century to jump-start the schooling of older children.

Beyond cultural differences that influence the "initial" level of education and human capital, technological and geographic differences play a role in causing differences in the size of γ. *Technological or geographic features* that cause γ to be high, increase the likelihood of a trap. In colder northern regions, farming is more physically demanding (hay, wheat, and dairy farming). This makes the relative productivity of children in agriculture low. As the Industrial Revolution began in England, there was a decline in heavy agriculture and an expansion in the "cottage industry," where textiles were produced at home. This shift in production raised the relative productivity of children significantly. The cottage industry was a relatively unique feature of early growth in England and this may be one reason why the education of older children, whose productivity was relatively high, lagged other countries without a prominent cottage industry.

In southern climates, the farming is less heavy (sugar, cotton, and rice) and the relatively productivity of children is higher. This is one explanation for why education in the southern U.S. lagged that of the northern U.S.. Geographic differences that lead to schooling differences are also consistent with more general statistical findings using cross-country data. Warmer climates are correlated with lower income among developing countries. There are many possible explanations for this finding, but one could be that children have higher relative productivity in the farming of warmer climates and as a result receive less schooling. High levels of schooling, and the resulting economic development, allow countries to become less dependent on their geography. Geography, however, can create poverty traps for developing countries, especially those in warmer climates where children are relatively productive and the opportunity cost of educating older children is high.

4.5.1 Escaping the Poverty Trap

How can an economy escape the trap? Schooling is low in a poor country because the value of forgone earnings associated with sending older children to school is a relatively large fraction of parent's income. The value of forgone earnings is high because households have many children and because parental earnings are low. The poverty trap can be removed if parental earnings are increased relative to the earnings of older children. This would make it more costly to have many children (because of the lost work time and forgone consumption of parents associated with child rearing) and it would lower the relative value of children's work in total family income.

What is needed is more schooling of *older* children, so that when they become parents their earnings (based on $e_t > \bar{e}$) exceed the earnings of their older children (based only on $\bar{e}$)—thereby making children more costly and relatively less important in generating family income as time goes on.

One could try imposing *compulsory schooling or child labor laws* that mandate less work and more schooling for older children. The problem with this approach is that parents are made worse off by the mandate and thus will evade the law. Because many poor households of developing countries live in remote rural areas, it is expensive for authorities to monitor household behavior to see if the laws are being followed.

A better approach to escaping the schooling trap is Mexico's *Progresa* program. In 1997 Mexico began *Progresa*, a program designed to increase human capital in poor families by paying families to send their children to school and to visit health care providers. Grants and subsidies are provided directly by the government to the mothers of older children who attend school. The grants cover about 2/3 of what the child would receive in full time work. This reduces the marginal cost of schooling from $w_t\gamma\bar{h}$ to $(1-2/3)w_t\gamma\bar{h} = w_t\gamma\bar{h}/3$ which is equivalent to lowering the relative productivity of child labor to just 1/3 of its original value. We have seen that an increase in γ shifts the schooling equation down and increases the likelihood of a schooling trap (see the change from Fig. 4.4 to Fig. 4.3). The Progresa policy does the opposite—it effectively reduces γ and shifts the schooling equation up, reducing the likelihood of a schooling trap. Chapter 5 examines the Progresa policy in a complete growth model.

4.6 The Malthusian Era

G7 is one of the more dramatic growth facts: the world experienced no significant increase in living standards before 1800. The best theory explaining this remarkable fact was provided by Thomas Malthus, a classical economist thinking about the perpetual stagnation of per capita income right around 1800. Malthus is sometimes criticized for coming up with his pessimistic theory of stagnation at the dawn of the

Industrial Revolution and the start of modern economic progress. The criticism is unfair. Malthus's simple theory elegantly explains the absence of economic growth that characterizes the vast majority of human existence and continues to be relevant to many developing countries today. He can be blamed a bit for "extrapolating from past data without due modesty" (Buchholz (1989, p. 56)) but the Industrial Revolution that pushed the world out of the Malthusian Era is difficult to explain in hindsight, let alone predict in 1800.

Malthus's theory works in a *traditional economy* where land, and not physical capital, is the key input that complements work effort in the production of goods. Traditional economies before 1800 experienced plenty of inventions and ideas that could have lifted productivity and living standards: domestication of animals, irrigation systems, axes and saws, wagons, plows, spinning wheels, gunpowder, rifles, and many more. Malthus saw population growth as the main reason that innovations failed to raise living standards in the long-run. Whenever workers became more productive, they lived longer and had more children. The increase in longevity and fertility increased the population and crowded the available land and natural resources. The lower land to labor ratio reduced worker productivity and negated the effects of the innovations on production per person.

The inverse relationship between population size and productivity, predicted by Malthus, is displayed in Fig. 4.5. The figure is similar to one found in Galor (2011). Oded Galor's work has revitalized interest in Malthus's emphasis on population growth as a factor limiting advances in living standards. The impressive historical

Fig. 4.5 Farm real wage and population data from England. (*Source*: Clark (2007) and Wrigley et al. (1997))

data in the figure, going all the way back to 1250, was recently collected and organized by Gregory Clark (2007). Figure 4.5 shows that increases in worker productivity and real wages were associated with increases in population that pulled wages back down as described by Malthus.

In this section we construct a traditional economy that captures the key features of the world that Malthus observed. Combined with our theory of fertility from previous sections we can reproduce the population dynamics that Malthus had in mind. In later chapters, the traditional sector developed here is combined with the modern sector modelled in Chaps. 2 and 3 to form a *two-sector* analysis.

4.6.1 A Traditional Economy

Three new assumptions are used to characterize a traditional economy. First, the focus shifts from the use of physical capital to the use of land and other natural resources in producing goods. Traditional producers use tools but they do not have the same impact on production as the large plant and equipment used in the factories of a modern economy. Second, the economy is in a schooling trap where only children that are too young to work receive any formal education. One can think of θ as being too low or γ as being too high to motivate schooling of older children prior to the Industrial Revolution. Third, because output is low, the idea that people need a minimum level of consumption to survive is made explicit.

The production function of the traditional economy is

$$Y_t = AL^{\alpha}H_t^{1-\alpha}, \tag{4.9}$$

where a fixed quantity of land, L, takes the place of manmade physical capital as an input in production. The labor productivity index is assumed to be constant and set equal to one, $D_t \equiv 1$. While there were certainly innovations in production before 1800, they were not as regular and ongoing as after the Industrial Revolution. The more sporadic inventions of the Malthusian Era are captured by irregular discrete changes in A. Human capital achieved from early education is also set equal to one, $\bar{h} \equiv 1$, so that $H_t = N_t$. For simplicity only, in this section we assume that children do not work. If we divide both sides of (4.9) by N_t, we can write worker productivity as

$$y_t = A(L/N_t)^{\alpha} \tag{4.9$'$}$$

We continue to assume that competitive markets for labor and goods exist. There is no market for land. Land is simply passed down from parents to their children. The existence of land markets is discussed in detail in later chapters. A competitive labor market means that workers are paid a wage equal to the marginal product of labor,

$$w_t = (1-\alpha)Al_t^{\alpha}, \tag{4.10}$$

where l is the land to labor ratio chosen by the landowner hiring labor. The landowner receives land rents that equal the difference in output and labor costs. In the economy as a whole, land rents are

$$R_t^L L = Y_t - w_t N_t = N_t\left(Al_t^\alpha - w_t\right) = N_t \alpha A l_t^\alpha. \tag{4.11}$$

Equation (4.11) implies land rents per unit of land are

$$R_t^L = \alpha A l_t^{\alpha-1}. \tag{4.12}$$

Household saving does not play the crucial role in building up the physical capital stock that it did in the model of a modern economy, so we can simplify things on the household side by assuming everyone lives for just *one* period. Household preferences are then

$$U_t = \ln\left(c_t - c\right) + \psi \ln\left(n_{t+1} w_{t+1} h_{t+1}\right) = \ln\left(c_t - c\right) + \psi \ln\left(n_{t+1} w_{t+1}\right), \tag{4.13}$$

where c is the constant level of subsistence consumption needed for survival. Only consumption above c generates utility.

Each household earns the wage, w_t. Households are identical, so each household owns $L/N_t \equiv l_t$ units of land and receives land rents equal to $R_t^L l_t = \alpha A l_t^\alpha$. Total household income is then $w_t + R_t^L l_t = y_t$. There is a loss in adult consumption of $\overline{\eta} y_t$ to raise each child.

Households choose c_t and n_{t+1} to maximize (4.13) subject to the household budget constraint

$$c_t + \overline{\eta} y_t n_{t+1} = y_t. \tag{4.14}$$

The optimal choices of consumption and fertility are (see the chapter Appendix)

$$c_{1t} = \frac{1 + \psi(c/y_t)}{1 + \psi} y_t \tag{4.15a}$$

$$n_{t+1} = \frac{\psi}{1 + \psi} \frac{1 - (c/y_t)}{\overline{\eta}}. \tag{4.15b}$$

From (4.15) we see if y_t were to increase over time the marginal propensity to consume out of income would fall and converge to the constant $1/(1 + \psi)$ and fertility would rise, approaching the constant value $\psi/\overline{\eta}(1 + \psi)$. The rise in fertility associated with a rise in income offers a possible explanation for the first part of *G3*. However, this does not happen in a Malthusian world because the rise in n_{t+1} raises $N_{t+1} = n_{t+1} N_t$, lowering both $l_{t+1} = L/N_{t+1}$ and $y_{t+1} = Al_{t+1}^\alpha$. Prosperity increases fertility and crowds the available land, creating a negative feedback that limits any gain in prosperity. Something has to break this connection before income can rise persistently.

4.6.2 Malthusian Fertility-Income Dynamics

We focus now on the dynamics associated with fertility. We want to answer the question, "How great is the tendency for fertility to limit growth?"

First, imagine the events qualitatively. Suppose in period t an innovation occurs that causes a jump in TFP to a value $A' > A$. In period t there is an increase in y_t and by (4.15) both consumption and fertility. Next period, there is a higher population $N_{t+1} = n_{t+1}N_t$ that reduces land holdings per worker and reduces at least some of the positive effect of the rise in TFP, forcing y_{t+1} back toward the original pre-innovation value of y_t. To see what happens at this point we need the help of a useful diagram that illustrates the dynamics in the same way that transition equations help track the paths of physical and human capital.

The diagram involves two distinct figures. The first figure is simply a sketch of the traditional production in per capita terms, $y_t = A(L/N_t)^{\alpha}$. For a given value of A, there is an inverse relationship between y_t and N_t because higher values of N_t reduce the land-labor ratio and y_t. The inverse relationship between y_t on the horizontal axis and N_t on the vertical axis is sketched in Fig. 4.6.

Next we need to track population changes using (4.15b). Higher values of y_t increase n_{t+1} and N_{t+1}. Note that (4.15b) tells us there is a value of y_t, call it $\bar{y}$, that causes the household to have exactly one child defined by

$$n_{t+1} = 1 = \frac{\psi}{1+\psi} \frac{1 - \left(c/\bar{y}\right)}{\bar{\eta}}.$$

This is important because in terms of population dynamics $\bar{y}$ defines a constant population size, i.e. a *population steady state* with $n_{t+1} = 1$ and $N_t = N_{t+1}$. Values of

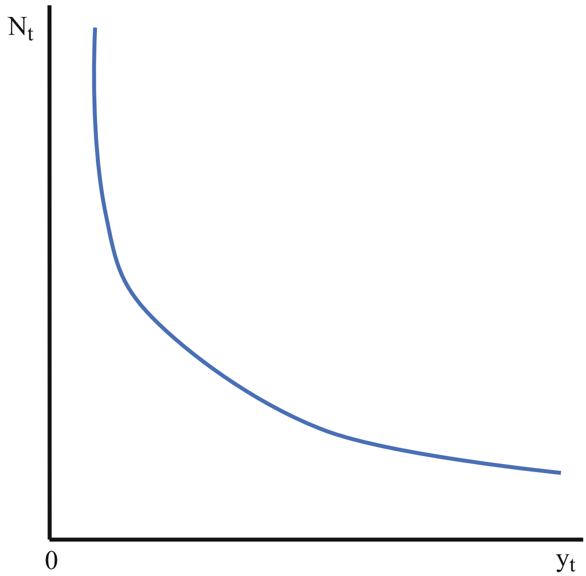
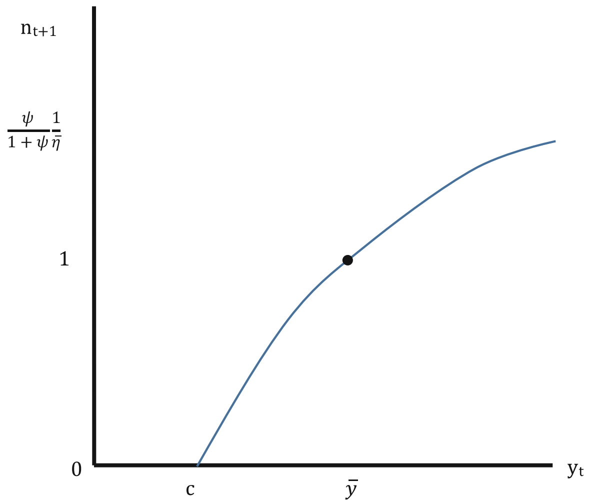


Fig. 4.6 Worker productivity

Fig. 4.7 Fertility



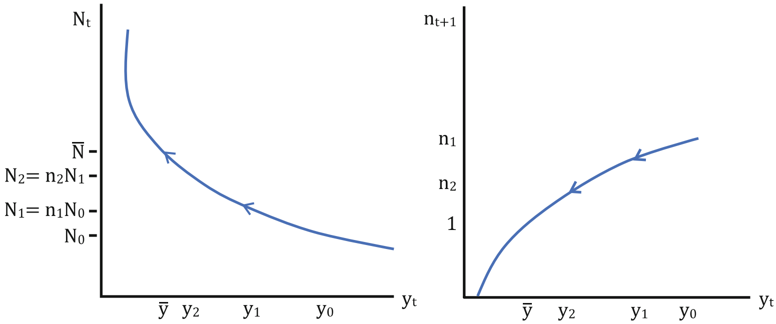


Fig. 4.8 The Malthusian model

y below $\bar{y}$ cause such low fertility that households are not replacing themselves and the population shrinks. Values of y above $\bar{y}$ cause households to have more than one child, making population grow over time.

We sketch the positive relationship between y_t and n_{t+1} with y_t on the horizontal axis and n_{t+1} on the vertical axis in Fig. 4.7. We have already established that if $y_t = \bar{y}$ then $n_{t+1} = 1$ and the population does not change over time. As y_t increases, (4.15b) tells us n_{t+1} approaches its maximum value of $\frac{\psi}{1+\psi}\frac{1}{\bar{\eta}}$. As y_t falls toward subsistence consumption, n_{t+1} falls to zero.

Now combine Figs. 4.6 and 4.7 to form Fig. 4.8. Suppose we start at the point $y_t = \bar{y}$ with $n_{t+1} = 1$. Here the population size that starts at $\bar{N}$ will remain at $\bar{N}$ over time. If instead, we start at $y_0 > \bar{y}$ and $N_0 < \bar{N}$ in period 0, then $n_1 > 1$ and $N_1 > N_0$. The greater population in period 1 means $l_1 < l_0$ and $\bar{y} < y_1 < y_0$, we have moved back and up the worker productivity curve as the population grows and living standards fall.

With $\bar{y} < y_1$, we know the dynamics is not over because n_2 remains above 1 and the population continues to increase toward $\bar{N}$. The dynamic process continues until the economy converges to the steady state $\bar{y}$, $\bar{N}$.

Malthusian Dynamics

$$y_0 \rightarrow n_1 \rightarrow N_1 = n_1 N_0 \rightarrow y_1 \rightarrow n_2 \rightarrow N_2 = n_2 N_1 \rightarrow y_2 \rightarrow \cdots \rightarrow \bar{y} \rightarrow \bar{n} = 1 \rightarrow \bar{N}$$

If we instead started with $y_0 < \bar{y}$ and $N_0 > \bar{N}$, we would trace through the dynamics, this time with y rising and N falling over time, again converging to the steady state $\bar{y}$, $\bar{N}$. Thus the steady state is *dynamically stable*—no matter where you begin, the economy tracks to the steady state equilibrium.

Now back to the main question,

To what extent does population change limit economic progress?

Suppose we start in the steady state depicted in Fig. 4.8. Imagine an innovation in period 0 that causes a jump in TFP to a value $A' > A$. This changes the relationship between y_t and N_t because higher values of y_t are now possible for any given value of N_t, i.e. the worker productivity curve has shifted out to the right as depicted in Fig. 4.9. The value of y_0 jumps above $\bar{y}$ causing fertility to exceed one, $n_1 > 1$. The increase in TFP has (i) created a new steady state, at the point $\bar{y}$, $\bar{N}' > \bar{N}$, and (ii) moved the economy above the steady state value of worker productivity, $y_0 > \bar{y}$. We are in a similar position to the one we used in Fig. 4.8 to trace the dynamics of the economy. Fertility will exceed one and the population will increase, pushing the economy back to the same steady state $\bar{y}$ but now with the larger population, $\bar{N}' > \bar{N}$.

The answer to our main question is that population growth *completely eliminates* the potential economic progress associated with innovations. For a while there is greater prosperity but this causes fertility to rise. The population grows, lowering land per worker and worker productivity until the population stops growing. The

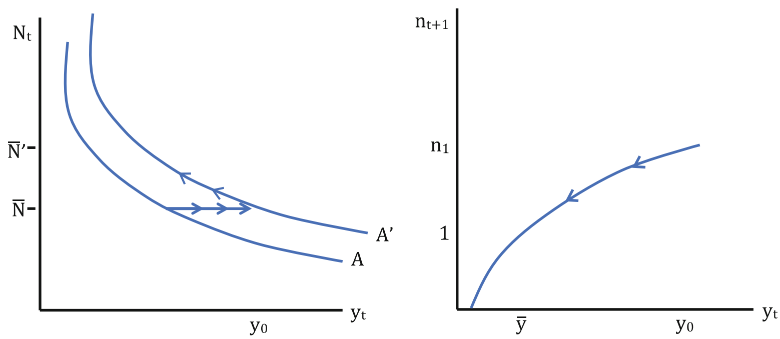


Fig. 4.9 Positive productivity shocks

population will only stop growing when worker productivity and income is pushed all the way back to $\bar{y}$.

The data in Fig. 4.5 has been explained. Figure 4.6 explains the inverse relationship between population and wages. Figure 4.7 explains why wages tend to gravitate toward a constant value. Wages rise temporarily causing population growth that reduces the wage back down toward the original value. Despite innovations, there is no long-run upward trend in wages and living standards—the *Malthusian Trap*.

4.6.3 Escaping the Malthusian Trap

Something changed around 1800 that allowed an escape from the Malthusian Trap, although not in all countries at the same time. Technological advances caused the first factories to appear that eventually pulled labor out of traditional production into firms of the modern sector (see Chap. 6 for a complete analysis). Associated with this fundamental structural transformation in the method and composition of production were several features that caused fertility to begin to fall, and not rise, with income. Factory wages replaced land rents as the major source of household income. As explained in Chap. 7, this change in the source of income raised the opportunity cost of raising children and lowered fertility.

The schooling poverty trap discussed above was also broken. Part of the reason was a rise in the return to schooling, as the technological advances increased the rewards for literacy, numeracy, and the flexibility needed to keep up with a more dynamic economy (Galor (2011, Chapter 4)). In our model, we can think of this as an increase in θ. A second reason was the onset of child labor and compulsory schooling laws that lowered the relative productivity of children (Doepke and Zilliboti (2005)), a decline in γ. The drop in fertility and the new reliance on manmade physical capital allowed for sustained increases in per capita incomes for the first time in human history.

4.7 Rising Fertility in Early Development

Our focus has been on explaining why fertility falls with development. However, fertility often rises in the early stages of development before it begins to fall (Galor and Weil (2000) and Jones (2001)). For example, in the early stages of the Industrial Revolution in England, fertility first rose modestly from a TFR of about 4.5 in 1750 to a TFR of about 5.5 in 1820 (Woods (2000)). Fertility began to fall after 1820, reaching a low of a little over 2 by 1940.

One possible explanation for the rise in fertility is a rise in the relative productivity of children as economies shift from agricultural production to informal manufacturing production in the traditional sector and early factories in the modern sector. The share of employment in England's agricultural sector fell during the eighteenth century, but it was largely offset by increases in family production outside of agriculture, as well as some expansion in early factory employment. As noted by Sokoloff and Dollar (1997, p. 289),

> Cottage manufacture (or putting-out), where workers labored at home as individuals or in family groups for piece rates, was common in England into the late nineteenth century. It was rare in the United States, however, where the overwhelming share of manufactures intended for sale came instead from centralized plants, which operated as manufactories or so-called nonmechanized factories. This mode of manufacturing organization, where workers routinely left home each day to labor together in a structure intended for that purpose, was also used in England but appears to have been much less prevalent.

Thus, from 1740 to 1820 there was a decline in agriculture, but an expansion in another aspect of the traditional sector, the "cottage industry." This caused a decline in employment opportunities for children in agriculture, but an expansion in both the cottage industry and in early factories of the modern sector (Cunningham (1990), Horrell and Humphries (1995), and Hudson and King (2000)). The earnings of children relative to adults were significantly higher in the factories, and especially the cottage industry, than in agriculture (Horrell and Humphries (1995, Table 5)). Thus, as the employment of children shifted from agriculture to domestic and formal industry, the relative productivity of children rose. In our model this would be captured by a general rise in γ from 1740 to 1820. A rise in γ lowers the cost of children and increases fertility. Fertility began to fall when γ fell back after the cottage industry declined and the employment opportunities for children in factories and mines were curtailed by child labor laws.

Galor and Weil (2000) suggest another possible explanation for the early rise in fertility, one related to the subsistence constraint. In the early stages of growth, before mandatory schooling and child labor laws, schooling may not rise above the education of young children. In the Malthusian model the importance of the subsistence constraint, c/y_t, falls with a rise with income. As seen in (4.15b), a rise in income lowers the value of forgone consumption associated with raising children and increases fertility via an income effect. Thus, the more regular technological advances associated with the Industrial Revolution lifted income and caused fertility to rise in the early stages of growth before schooling began to increase.

4.8 The Baby Boom

One cannot look at Fig. 4.2 without noticing the upward blip in the fertility data that temporarily interrupted the long-run downward trend in fertility. The rise in fertility between 1945 and 1965 is known as the Baby Boom. The Baby Boom has had long term consequences for the United States and other developed countries because it has contributed to the aging of their populations. The Baby Boomers have retired from the work force or are about to, accelerating a general downward trend in the number of workers per retiree. The downward trend in the worker to retire ratio is due to more than just the Baby Boomers retiring. As economies have prospered over the last two centuries, fertility has fallen and longevity has increased. These long-run demographic trends have led to an increase in the average age of the population.

Currently in the United States, there are 4.8 workers per retiree. As the century unfolds the ratio will fall to 2.8—instead of 5 workers to support each retiree, there will only be 3 (Kotlikoff and Burns (2012)). The change in the age composition of the country is particularly important because the benefits delivered via retirement programs, Social Security and Medicare, have become quite generous. Taxes on the working population to support the retirement programs will have to rise significantly over this century.

What caused the Baby Boom? One theory, offered by Doepke et al. (2015), argues the following economic factors were at play. During World War II there were labor shortages due to the heavy mobilization of troops. The labor market shortages pulled older women out of the home and into the workforce; women in their forties and fifties who had already raised their children. These women largely decided to remain working after the war ended, depressing the work opportunities for young women who typically worked for some years before marrying. Left with depressed work and wage opportunities, the young women decided to marry and start their families earlier. The weak labor market for young women can be interpreted as a temporary decline in η, lowering the opportunity cost of raising children. These conditions persisted until the older generation of working women began to retire, which is why the Baby Boom lasted for 20 years.

4.9 One-Child Policy

Countries have long understood the negative effects of population growth on living standards and have attempted to encourage a reduction in fertility in various ways. In the late 1970s, China went beyond encouragement by enacting a law prohibiting families, except in special circumstances, from having more than one child. At this time, the fertility rate had already been falling dramatically but it was still high, just below three children per woman. Thus, the policy clearly imposed a binding constraint on household behavior. The purpose of the policy was to prevent population crowding and reduce unemployment and the demand for natural resources. Recently, China has relaxed its policy to allow two children per household.

In this section we think about how an exogenously imposed reduction in fertility would affect the equilibrium in our model. Obviously, constraining household choice in any manner will reduce welfare for the initial generation of households in both sectors—adding any type of constraint on choice can only reduce possibilities and welfare. However, there are some less obvious impacts.

One possible benefit to the one-child policy is to increase schooling. We find that this is the case in our model. To begin, one can combine the necessary conditions for optimal schooling and household consumption to get the following expression

$$e_t = \frac{\psi\theta c_t}{n_{t+1} w_t \gamma \bar{h}}.$$

The numerator captures any wealth effect on schooling that is associated with a rise in consumption. The denominator is the marginal cost of schooling measured in terms of the forgone wages that children could earn working. In the unconstrained model, when fertility is a choice variable, changes in family resources affect consumption and fertility choices proportionately, leaving schooling independent of fertility, as you see in (4.3b). In our model, schooling affects fertility with no reverse causation. Larger families, due to higher income or stronger preferences for children, are not associated with lower schooling per child. The quantity-quality trade-off in our model is driven by rising schooling that causes the cost of children to increase and fertility to fall.

Now suppose fertility is *exogenously* determined. In this case, one can show that higher fertility both reduces family resources and raises the cost of schooling, causing a decline in schooling (see *Problem* 16). Thus, an exogenous decline in fertility increases schooling.

Using data from China, recent empirical studies have attempted to estimate the effect of the one-child policy on schooling. The evidence thus far is mixed. For example, Rosenzweig and Zhang (2009) provide estimates supporting the idea that restricting the number of children *increases* schooling per child. On the other hand, Qian (2013) finds that restricting the birth of a second child, for families with only one child, actually causes the schooling of the first child to *fall* (although she finds the effect of restricting further children, in families with more than one child, increases schooling per child).

4.10 Human Capital and Inequality

We have seen that human capital accumulation raises worker productivity and living standards. Human capital also plays an important role in determining wage inequality over the course of development.

Until recently, economists have been comfortable assuming that technological progress can continue indefinitely without being subject to the phenomenon of diminishing returns. Technological progress continually raises the demand for human capital. This can be seen explicitly by recalling that the rental rate firms are willing to pay to a unit of effective labor supply,

$$w_t D_t = (1-\alpha) A E_t g_t^{\mu(1-\alpha)} k_t^{\alpha},$$

is increasing in the variable E_t that captures productivity effects from technological change.

More sophisticated models treat skilled and unskilled labor as distinct complementary inputs in production. In this case there is also the possibility of *skill-biased* technological progress that raises the *relative* demand for skilled labor. Rather than technical innovations proportionally raising the productivity of human capital in general, regardless of its level (as in our model), there may be innovations that raise the productivity of high-human capital workers more than low-human capital workers. In our model, more educated workers receive higher wages than less educated workers if the education gap widens, but not because of technological progress because the market rental rate on a given unit of human capital is common to all workers. Over the second half of the twentieth century, *both* a widening education gap across workers *and* skill-bias technological change have increased the relative wage of the highly educated worker.

Research by Claudia Goldin and Lawrence Katz (2008) quantifies how much skill-biased technological progress raised the relative demand for high-skilled labor over the Post WWII period in the United States. They find that the *demand* for high skilled labor grew at an approximately constant rate over the period. Changes in the relative wage paid to high-skilled labor were caused by the degree to which the *supply* of skilled labor was able, or unable, to keep pace with the ongoing demand.

From 1950 to 1980 the supply of college graduates increased at about the same pace as the demand for college-educated labor, leaving the relative wage, or *skill premium*, paid to highly educated labor approximately unchanged. Think of the relative demand curve for and the relative supply curve of skilled labor shifting out by equal amounts, leaving the relative market wage paid to skilled labor unchanged (the absolute amount was rising but at the same rate as for workers with less education).

However, as we have previously discussed, the ability of a country to increase its supply of human capital is eventually subject to diminishing returns. After 1980, increases in the average years of schooling began to slow as the percentage of young workers receiving college degrees stagnated. The demand for skilled-labor began to outpace the supply of skilled labor causing the relative wage paid to highly educated workers to increase. From 1980 to 2005 the skill premium for a college graduate more than doubled. In 1980 a college graduate earned 37% more than a high school graduate. The skill premium rose to 87% by 2005 (Goldin and Katz (2008, p. 95)).

Much of the gain in the relative wages of college graduates is concentrated among the relatively small percentage of the workforce with *graduate* degrees. In the U.S., from 1980 to 2012, full-time male workers with a *graduate* degree saw their real earning rise 1.1% annually. For college graduates with a bachelor's degree, the rise in real earnings was a paltry 0.5%. Those with some college saw no gain in real wages and those with high school degrees currently receive lower real earnings than they did in 1980. Only 30% of workers eventually complete a four-year college degree, so the vast majority of workers have not experienced a rise in real earnings since 1980 (Autor (2014)). Most workers have a high school degree or less. The combination of a growing education gap and a rising skill premium has generated a dramatic rise in wage inequality in the United States.

The role of human capital in creating rising inequality in the later stages of development is not unique to the United States. Wage inequality has been on the rise in most developed countries over the last 35 years (Cingano (2014)). The United States has a particularly high degree of income inequality, but inequality is predicted to continue its rise in advanced countries generally. Trends in inequality suggest that the average OECD country will reach the current level of income inequality in the United States by mid-century (OECD (2014)). Rising wage inequality in developed economies is achieving the status of a stylized fact (a *G13*?).

4.11 Exercises

Questions

1. Write out the expressions for the wages of an adult and the wages of an older child. Identify three reasons why the wages of adults might be higher than wages of older children.
2. Describe the theory of schooling and fertility in words.
3. Explain how schooling and fertility are affected by (a) parent's income and (b) the relative productivity of child labor.
4. Explain how fertility is affected by the years of schooling children receive.
5. Why are the schooling and fertility traps linked?
6. Explain why, if the parent's level of schooling is sufficiently high, fertility will fall over time.
7. How does an increase in the market rental rate for human capital, w_t, affect schooling and fertility?
8. Why are the determinants of cross-sectional variations in fertility within a country different than time series variation in the country's average fertility?
9. Use a diagram of the schooling equation to explain why a higher value for γ increases the likelihood of a schooling-trap.
10. Use the diagram of the schooling equation to show how if the parent's level of schooling is sufficiently high, the level of schooling will rise over time.
11. Discuss the ways that (a) culture (b) technology and (c) geography can affect whether a country finds itself in a schooling-trap.
12. Discuss the effectiveness of using the following polices to escape the schooling trap.
    (a) compulsory schooling laws
    (b) child labor laws
    (c) Progesa-style family subsidies
13. What is the Malthusian Era?
14. How does a traditional economy differ from a modern economy?
15. What determines fertility in a traditional economy?
16. Intuitively describe the Malthusian model—i.e. the two key relationships underlying Fig. 4.8.
17. Carefully explain the reasons for the slopes of the figures sketched in Figs 4.6 and 4.7. How do changes in the following parameters shift each curve?
    (a) c
    (b) L
    (c) A

18. Use Fig. 4.9 to explain the Malthusian theory of stagnant living standards.
19. The data in Fig. 4.5 was influenced by the Black Plague (also known as the Black Death and the Bubonic Plague) that swept through Europe in the mid-1300s. Analyze the economic effects of a discrete decline in population using the Malthusian model.
20. Why might fertility initially rise as economies develop?
21. What is the Baby Boom? What may have caused it?
22. Discuss how an *exogenous* reduction in fertility, such as those under the One-Child Policy in China, affects schooling per child. Explain why endogenous variations in family size, due to variations in parental income or preferences across households, have no effect on schooling per child. What then creates the quantity-quality trade-off?
23. Why has wage inequality tended to rise as developing economies mature? What evidence supports your explanation?
24. Use the theories of this chapter to explain growth facts *G2*, *G3*, and *G7*.

Problems

1. Derive the optimal behavior for fertility, schooling, and saving given in (4.3). What is the causal connection between fertility and schooling? What condition must hold for there to be a poverty trap where parents and their children receive the minimum level of schooling?
2. Now suppose that learning of older children requires the family to purchase goods inputs as was discussed in Chap. 2 (continue to assume that learning of young children is only a function of their time). Let the quantity of goods inputs purchased per child be denoted by x_t and the price to the family of purchasing one unit of the goods inputs by p_t. Write the human capital production function as $h_{t+1} = x_t^{\theta_1} e_t^{\theta_2}$. Introduce the schooling expenditures, $p_t\, x_t$, into the family's budget and derive the new expression for the choice of schooling time solely as a function of parent's schooling time and exogenous variables, as in (4.3b). Discuss the new determinants of the poverty trap and how policy can be used to generate a human capital take-off.
3. Extend your analysis in *Question* 2 to include fertility. Derive the new expression for fertility and discuss how the goods cost of schooling affects the level of fertility.
4. Think more broadly about human capital inputs to include health and nutritional investment in children. Apart from the good inputs used in education, what alternative interpretations can then be given to x_t? Does the relationship between x_t and e_t make sense under the new interpretation? How does the introduction of x_t help to further explain the income gaps across countries?

   Problems 5–8 use the simplified schooling transition equation given by (4.5).
5. Suppose the schooling equation crosses the 45-degree line exactly at the point $e_t = e_{t-1} = \bar{e} = 1/9 = 0.1111$, as in Fig. 4.4 At that point, the schooling equation must satisfy $\bar{e} = \frac{\eta}{\gamma} - 0.50 = 0.1111$ (Why?).

   (a) For this to be true, what value must $\frac{\eta}{\gamma}$ have?
   (b) As mentioned in the text, empirical studies suggest that $\gamma = 0.30$ is a reasonable estimate. If we assume this value for γ, what value must η to be consistent with your answer to part (a)?
   (c) Write the schooling equation for the parameter values set as above.
   (d) If parents start with an a value of e equal to 0.12, use the schooling equation to compute the dynamic path of schooling until the maximum schooling is reached.
   (e) Use a sketch of the schooling equation, as in Fig. 4.4, to represent the dynamic path of schooling computed in (d).
6. Suppose that $\gamma = 0.40$ and $\eta = 0.20$. These parameter values create a schooling equation similar to the one depicted in Fig. 4.3.
   (a) Write out the schooling equation for these parameter values.
   (b) If parents start with a value of e equal to 0.20, use the schooling equation to compute the dynamic path of schooling.
   (c) How high must parent's initial schooling be before the economy escapes the schooling poverty trap? This requires that you use trial and error to find a value of e that creates a steady state associated with point B, say $\widehat{e}$, as depicted in Fig. 4.3.
   (d) Pick a value for parents' initial schooling slightly above your answer to (c) and compute the dynamic path until the maximum schooling is reached. Use an initial value of e equal to $\widehat{e} + 0.05$.
   (e) Use a sketch of the schooling equation, as in Fig. 4.3, to represent the dynamics you uncovered in (b)-(d).
7. For the parameter calibration we used in Sect. 4.4, what value of the parameter ψ is needed to make the number of children per parent exactly one in the situation with full education and no child labor? For this parameter value, what is the number of children per parent in a country stuck in the schooling-fertility trap?
8. Assuming the parameter values from Problem 5, suppose that the parent's initial education is 0.12. Compute the paths for schooling and fertility until the maximum schooling level is reached.
9. How would the introduction of a Progesa-style family subsidy affect the schooling equation in Fig. 4.4?
10. Show that $y_t = Al_t^\alpha = w_t + R_t^L l_t$.
11. Fill in the algebraic details involved in establishing the equalities in (4.11), and then in going from (4.11) to (4.12).
12. Maximize (4.13) subject to (4.14) to get (4.15).
13. What is the maximum potential value for fertility in the Malthusian model? What is the steady state value?
14. Use (4.15b) to find an expression for $\bar{y}$ in terms of the parameters of the model.
15. Use Fig. 4.8 to demonstrate that the Malthusian model is dynamically stable—no matter where the economy's initial value of y, it converges to $\bar{y}$.
16. Suppose a fertility limit is imposed by the government. Derive the schooling equation under the policy, assuming the imposed fertility limit is binding on household choice. Show how a reduction in the fertility limit affects schooling.

Appendices

Maximizing Utility with Fertility and Schooling

In addition to the standard necessary conditions for optimal life-cycle consumption from Chap. 2, the choices of n_{t+1} and e_t associated with maximizing (4.1) subject to (4.2), yield the following first order conditions

$$\frac{\psi\theta}{e_t} \leq \lambda_t n_{t+1} w_t D_t \gamma \bar{h}$$

$$\frac{\psi}{n_{t+1}} = \lambda_t[\eta w_t D_t h_t - (T - e_t) w_t D_t \gamma \bar{h}],$$

where λ_t is the Lagrange multiplier.

The first equation says the marginal utility of additional child quality, measured by the child's adult human capital, must be equated to the marginal value of consumption lost from allowing children of working age to attend school. The strict inequality holds when the marginal cost of educating children beyond the schooling received in their early years, $\bar{e}$, exceeds the marginal benefit. In this case, parents are content to set $e_t = \bar{e}$, i.e. to have their children begin work as soon as they are able.

The second equation says the marginal utility of additional children must be equated to the marginal value of lost consumption associated with raising a child. Consumption is lost from having an additional child because we assume the cost of children exceeds the earnings that older children bring to the household (otherwise parents would always choose the, biologically determined, maximum number of children).

Fertility in the Traditional Economy

In the traditional economy from Sect. 4.6, the household maximizes the utility function $U_t = \ln(c_t - c) + \psi \ln(n_{t+1} w_{t+1})$ subject to the budget constraint $c_t + \overline{\eta} y_t n_{t+1} = y_t$. The first order necessary conditions for the optimal choices of c_t and n_{t+1} are

$$\frac{1}{c_t - c} = \lambda_t$$

$$\frac{\psi}{n_{t+1}} = \lambda_t \overline{\eta} y_t,$$

where λ_t is the Lagrange multiplier. The two first order conditions imply $\overline{\eta} y_t n_{t+1} = \psi(c_t - c)$. Substituting this expression into the budget constraint and solving for c_t yields (4.15a). Substituting (4.15a) into the expression $\overline{\eta} y_t n_{t+1} = \psi(c_t - c)$, gives us (4.15b)

$$n_{t+1} = \frac{\psi}{1+\psi}\frac{y_t - c}{\overline{\eta} y_t} = \frac{\psi}{(1+\psi)\overline{\eta}}(1-(c/y_t)).$$

References

Autor, D., 2014, "Skills, Education, and the Rise in Earnings Inequality among the Other 99 percent," *Science*, 344, 843–851.

Banerjee, A., and Duflo, E., 2011, *Poor Economics*, Public Affairs, New York.

Becker, G, 1960, "An Economic Analysis of Fertility," in National Bureau of Economic Research, *Demographic and Economic Change in Developed Countries*, Princeton University Press, Princeton.

Becker, G., 1981, *A Treatise on the Family, Harvard University Press*, Cambridge.

Becker, G. and Lewis, H., 1973, "On the Interaction between the Quantity and Quality of Children," Journal of Political Economy, 81 (20, S279–S288.

Buchholz, T., 1989, *New Ideas from Dead Economists*, New York: Penguin Books.

Cingano, F., 2014, "Trends in Income Inequality and its Impact on Economic Growth," OECD Social, Employment, and Migration Working Papers, No.163, Paris: OECD Publishing.

Clark, G., 2007, "The Long March of History: Farm Wages, Population, and Economic Growth, England 1209–1869," *Economic History Review*, 60(1), 97–136.

Cunningham, H. 1990, The Employment and Unemployment of Children in England c. 1680–1851, *Past and Present* 126:115–150.

Doepke M. and Zilibotti, F, 2005, "The Macroeconomics of Child Labor Regulation," *American Economic Review* 95(5):1492–1524.

Doepke, M., Hazan, M., and Maoz, Y., 2015, "The Baby Boom and World War II: A Macroeconomic Analysis," *Review of Economic Studies*, 82(3), 1031–1073.

Fernihough, A., 2017, "Human Capital and the Quantity-Quality Tradeoff during the Demographic Transition," *Journal of Economic Growth*, 22, 35–65.

Galor O., 2011, *Unified Growth Theory*. Princeton University Press, Princeton, New Jersey.

Galor, O. and Moav, O., 2002, "Natural Selection and the Origins of Economic Growth," *Quarterly Journal of Economics*, 117 (4), 1133–1192.

Galor, O., and Weil, D., 2000, "Population, Technology, and Growth: From Malthusian Stagnation to Modern Economic Growth, *American Economic Review*, 90(4), 806–828.

Goldin, C., and Katz, L., 2008, *The Race between Education and Technology*, Cambridge: Harvard University Press.

Haines, M, 2000, "The Population of the United States, 1790–1920," in S. Engerman and R. Gallman (editors), *The Cambridge Economic History of the United States*, Cambridge University Press, New York.

Horrell, S. and Humphries, J., 1995, "The exploitation of Little Children: Child Labor and the Family Economy in the Industrial Revolution," *Explorations in Economic History*, 32 (4):485–516.

Hudson, P. and King, S., 2000, "Two Textile Townships, c. 1660–1820: A Comparative Demographic Analysis," *Economic History Review*, 53(4):706–741.

Jones, C., 2001, "Was the Industrial Revolution Inevitable? Economic Growth over the Very Long Run," B.E. Journal of Macroeconomics 1(2), 1–45.

Kotlikoff, L., and Burns, S., 2012, *The Clash of Generations*, Cambridge, MA: MIT Press.

Lord, W. and Rangazas, P., 2006, "Fertility and Development: The Roles of Schooling and Family Production," *Journal of Economic Growth*, 11(3), 229–261.

OECD, 2014, *Policy Challenges for the Next Fifty Years*, Economic policy Paper No.9, Paris: OECD Publishing.

Qian, N., 2013, "Quantity-Quality and the One-Child Policy: The Only-Child Disadvantage in School Enrollment in Rural China," Yale University Working Paper, New Haven.

Rosenzweig, M., and Zhang, J., 2009, “Do Population Control Policies Induce More Human Capital Investment? Twins, Birth Weight, and China’s One Child Policy, *Review of Economic Studies*, 76 (3), 1149–1174.

Sokoloff, K., and Dollar, D., 1997, “Agricultural Seasonality and the Organization of Manufacturing in Early Industrial Economies: The Contrast between England and the United States,” *Journal of Economic History*, 57(2), 288–321.

Woods, R., 2000, “The Demography of Victorian England and Wales”. Camridge: Cambridge University Press.

Wrigley, E., Davies, R., Oeppen, J., and Schofield, R., 1997, *English Population History from Family Reconstitution 1580–1837*, Cambridge: Cambridge University Press

5 A Complete One-Sector Neoclassical Growth Model

In this chapter we put all the elements discussed in previous chapters together. The model includes private capital from Chap. 2, government capital and taxation from Chap. 3, and fertility and schooling from Chap. 4. The features are combined to study large income differences across rich and poor countries, what is known as *development economics*. At the dawn of the Industrial Revolution, the differences in per capita income across countries were relatively modest. Per capita income in the richest countries was only 2–4 times greater than per capita income in the poorest countries. Over the course of the last two centuries, per capita income of the rich countries has diverged from that of the poor countries. By the end of the twentieth century, rich countries were 20–40 times richer than poor countries. What could explain stylized growth fact *G4*, dramatically large gaps in living standards?

As we have seen in Chap. 2, the neoclassical model of physical capital accumulation was unable to explain the relatively modest seven-fold gap in per capita income in U.S. history between 1870 and 1990. It is clear that significant extensions to the neoclassical model are needed to explain the huge long-run income gaps that have evolved across countries in the past two centuries. In this chapter we explore how far the human capital, fertility, and government policy extensions will take us.

We begin by modeling a representative rich and poor country. One of the reasons that the poor country has low income stems from a *poverty trap* that keeps schooling low and fertility high. The second reason for its low income directly relates to fiscal policy. Many poor countries have high tax rates, high shares of government consumption, and little investment in public infrastructure. We calibrate the preferences of the government based on the fiscal policy outcomes that we observe across some rich and poor countries. Our calibration exercises reveal that the extensions from Chaps. 3 and 4 go a long way toward explaining large income gaps.

The large income gaps that appeared in the second half of the twentieth century motivated the rich countries of the world to provide various types of aid to poor countries with the hope of accelerating their growth, thereby allowing a convergence of living standards. One of the issues in providing aid is how the policies of the recipient country's government might respond. This is one of the reasons that we

S. Das et al., *Economic Growth and Development*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-319-89755-4_5

extend the model to make fiscal policy endogenous. It allows us to examine the factors that might cause the poor country's government to favor or oppose pro-growth policy reforms. In cases where the government clearly opposes a particular reform, we can also ask what the "aid cost" would be to convince the government to accept the reform.

5.1 A Theory of Income Differences

This section focuses on the sources of poverty. We create a model where a country is poor for two fundamental reasons. First, there is a poverty trap that keeps human capital low and fertility high. Second, the government of the poor country sets relatively high tax rates that are primarily used to finance government consumption rather than infrastructure investment.

5.1.1 Households

The theory of household behavior comes from Chaps. 2, 3 and 4. Households value consumption over the two periods of adulthood (c_{1t}, c_{2t+1}) and the adult earnings $((1-\tau_{t+1})w_{t+1}D_{t+1}h_{t+1})$ of all their children (n_{t+1}), similar to the "warm glow" preference for intergenerational transfers from Chap. 2. Earnings are the product of the after-tax market rental rate for skills $((1-\tau_{t+1})w_{t+1})$, the productivity index (D_{t+1}), and the embodied skills or human capital (h_{t+1}) of the worker. Formally, preferences are

$$U_t = \ln c_{1t} + \beta \ln c_{2t+1} + \psi \ln\left(n_{t+1}(1-\tau_{t+1})w_{t+1}D_{t+1}h_{t+1}\right)$$

where $0 < \beta < 1$ and $\psi > 0$ are preference parameters.[1] Adult human capital of the child is given by $h_{t+1} = e_t^{\theta}$, where e_t is the time the child spends in school and where $0 < \theta < 1$ is a parameter that gauges the effect of schooling on human capital accumulation. For simplicity we ignore purchased goods and services inputs used in schooling (e.g. tuition) and focus only on student time.

Adults inelastically supply one unit of labor when young and zero units when old. Children have an endowment of $T < 1$ units of time that they can use to attend school (e_t) or work $(T - e_t)$. Children have less than one unit of time to spend productively because early in childhood they are too young to either attend school or to work, and in the later years of childhood they do not have the mental or physical endurance to learn or work as long as an adult.

[1]Galor and Moav (2002) generalize this specification by allowing for a separate utility weight on the quantity and quality of children. They then go on to develop an evolutionary theory in which households raise the weight they placed on the quality of their children over the course of economic development. Using this more flexible specification would increase the ability of our model to fit the stylized growth facts.

While children may work as they become older, thereby providing income to the family, they are also expensive to care for and feed. In this chapter we assume that to raise each child requires a loss of adult consumption equal to a fixed fraction η of the adult's first period wages. This interpretation of the cost of children is somewhat different than the forgone wages associated with raising children. Both costs are relevant but we only assume one or the other for simplicity. See the chapter Appendix for a detailed discussion of the differences between the two types of costs.

We think of the children as being too young to work over the early part of their lives or alternatively that a minimum amount of schooling is needed for the child to be productive. Under either interpretation, each child invests at least $\bar{e}$ units of time into learning during the first portion of their childhood. This gives older children $\gamma \bar{h}_t = \gamma \bar{e}^{\theta}$ units of human capital that can be used in production during the later years of childhood, where $0 < \gamma < 1$ reflects the fact that children have less physical strength or experience than an adult. Thus, per hour of work, a person is more productive in adulthood than in childhood because of greater strength and experience ($1 > \gamma$) and possibly because of additional schooling ($e_t \geq \bar{e}$).

The household maximizes utility subject to the lifetime budget constraint,

$$c_{1t} + \frac{c_{2t+1}}{R_t} + n_{t+1}\eta w_t D_t h_t = w_t D_t h_t + n_{t+1} w_t D_t \gamma \bar{h}(T - e_t),$$

where $R_t = 1 + (1 - \tau_{t+1})r_{t+1} - \delta$, so both wages and rental rates are subject to taxation.

The new terms introduced by the extensions are the lost adult consumption associated with raising children ($n_{t+1}\eta(1 - \tau_t)w_t D_t h_t$) and the family income generated by child labor ($n_{t+1}(1 - \tau_t)w_t D_t \gamma \bar{h}(T - e_t)$). The net cost of raising children is the difference in these two expressions.

The demand functions for children, schooling, and the assets used to finance retirement consumption are

$$\begin{aligned} n_{t+1} &= \frac{\psi(1-\tau_t)w_t D_t h_t}{(1+\beta+\psi)(\eta(1-\tau_t)w_t D_t h_t - (1-\tau_t)w_t D_t (T-e_t)\gamma \bar{h})} \\ &= \frac{\psi}{(1+\beta+\psi)\left(\eta - \gamma(T-e_t)\left(\bar{e}/e_{t-1}\right)^{\theta}\right)} \end{aligned} \tag{5.1a}$$

$$\begin{aligned} e_t &= \max\left[\frac{\theta}{(1-\theta)} \frac{\eta(1-\tau_t)w_t D_t h_t - (1-\tau_t)w_t D_t T\gamma \bar{h}}{(1-\tau_t)w_t D_t \gamma \bar{h}}, \bar{e}\right] \\ &= \max\left[\frac{\theta\left(\eta\left(e_{t-1}/\bar{e}\right)^{\theta} - \gamma T\right)}{\gamma(1-\theta)}, \bar{e}\right] \end{aligned} \tag{5.1b}$$

$$s_t = \left[\frac{\beta}{1+\beta+\psi}\right](1-\tau_t)w_t D_t h_t. \tag{5.1c}$$

From (5.1a), we see that fertility is positively related to adult income (numerator) and negatively related to the net cost of children (denominator). Equation (5.1b) says schooling is positively related to the minimum net cost of children (the net cost when children work as soon as they can, found in the numerator) and negatively related to the forgone earnings associated with schooling (denominator). Schooling is high when children are expensive to raise and forgone earnings are low. However, if children are sufficiently cheap and forgone earnings are high, then fertility is high and parents want only the minimum schooling for each child. Equation (5.1c) gives a saving function similar to that derived in Chap. 2, but now the saving rate is inversely related to the taste for children parameter, ψ. More expenditures on children means less saving.

Note that for fertility to fall, there must be a rise in schooling. An increase in schooling raises the wages of parents relative to their children, which raises the net cost of children and lowers fertility. From (5.1a) and (5.1b), we see that the schooling-fertility dynamic is independent of the after tax rental rates on physical and human capital. A higher rental rate paid to human capital raises the earnings of both parents and children but does not affect *relative* earnings or the net cost of children. Thus, the evolution of schooling and fertility is unaffected by the determinants of human capital rental rates such as fiscal policy, technological change, and physical capital accumulation.

For schooling to begin to rise, the initial generation of parents must have human capital that is sufficiently higher than the human capital of their working-age children. If this condition holds, schooling will rise each period. As schooling rises, the effect of a given increment in schooling has a diminishing effect on human capital formation and wages. The transition equation given by (5.1b) thus will eventually exhibit the standard properties of neoclassical growth as human capital accumulates. However, there is no guarantee that the economy has the proper initial conditions to generate growth in human capital. If the economy's initial human capital investments are too low, it will be stuck in a *poverty trap* where schooling remains at $\bar{e}$ indefinitely.

5.1.2 Firms

Production takes place within standard neoclassical firms that combine physical capital and human capital to produce output from a Cobb-Douglas technology

$$Y_t = AK_t^{\alpha}(H_t)^{1-\alpha}, \tag{5.2}$$

where effective labor supply has been expanded to include endogenous human capital acquired through schooling, as well as child labor, $H_t = \widehat{h}_t N_t D_t$. The variable $\widehat{h}_t \equiv h_t + n_{t+1}\gamma\bar{h}(T - e_t)$ is the human capital supplied by the adult and children of a single household.

The productivity index, D, is now a function of disembodied technology, E, and government capital per adult worker, G/N, and is given by

$$D_t = E_t^{1-\mu}(G_t/N_t)^{\mu}, \tag{5.3}$$

where $0 < \mu < 1$ is a constant parameter. This specification captures the idea that public infrastructure raises the productivity of the private sector. We assume that E progresses at the exogenous rate q.

Firms operate in perfectly competitive factor and output markets. This implies the profit-maximizing factor mix must satisfy

$$r_t = \alpha A g_t^{\mu(1-\alpha)} k_t^{\alpha-1} \tag{5.4a}$$

$$w_t D_t = (1-\alpha) A E_t g_t^{\mu(1-\alpha)} k_t^{\alpha}, \tag{5.4b}$$

where the de-trended, for exogenous technical progress, values of public and private physical capital are defined as $g_t \equiv G_t/E_t N_t$, and $k_t \equiv K_t/E_t \widehat{h}_t N_t$.

5.1.3 Capital Market Equilibrium

The firm's demands for private physical and human capital are implicitly given by the profit maximizing conditions in (5.4). The supplies of private physical and human capital from the households are made available for firms to rent in the factor markets and are given by,

$$K_{t+1} = s_t N_t \tag{5.5a}$$

$$H_t = N_t \widehat{h}_t D_t. \tag{5.5b}$$

Substituting (5.1c), (5.4b), and (5.5b) into (5.5a), gives the equilibrium difference equation for physical-capital intensity,

$$k_{t+1} = \left[\frac{\beta}{1+\beta+\psi}\right] \frac{(1-\tau_t)(1-\alpha) A g_t^{\mu(1-\alpha)} k_t^{\alpha} h_t}{(1+q) n_{t+1} \widehat{h}_{t+1}}, \tag{5.6}$$

a transition equation that bears a resemblance to the ones derived in Chap. 2. However, there are four important differences. First, the wages for workers, that provide the basis for household saving, are now affected by the endogenous evolution of public, as well as private, capital. Second, the growth rate of the economy's effective labor supply is now affected by the endogenous accumulation of human capital. Third, income taxation lowers the after-tax wage, saving, and private capital accumulation. Finally, the population growth rate is now endogenous because it is determined by the household's fertility choice.

5.1.4 Government

We assume the government officials who determine fiscal policy are some fraction, ς, of the population of private households, N_t. Government officials value their own consumption (c_t^g) as well as the welfare of the representative citizen according to a single period utility function, $\ln c_t^g + \phi U_t$, where φ is a positive preference parameter that gauges the relative weight the government places on the welfare of private households, U_t. We assume that the current government also cares about the government as an on-going institution (i.e. they care about the future operations of the government and the welfare of future government officials) and the welfare of the country's future citizens. The preferences of the government are given by[2]

$$\sum_{t=0}^{\infty} \beta^t \left(\ln c_t^g + \phi U_t\right). \tag{5.7}$$

These complicated preferences make it explicit that the government's concerns extend indefinitely into the future. This is because there is no natural time horizon for government planning. Maximizing an objective function such as (5.7) is somewhat difficult but it turns out that the solutions for the optimal fiscal policy are surprisingly simple.

The government budget constraint is

$$c_t^g \varsigma N_t = \tau_t Y_t - I_t^g. \tag{5.8}$$

The left-hand side gives the government's consumption expenditures. The right-hand side is the difference between government tax revenue, net of transfers, and government investment in public capital. Public capital evolves according to the difference equation

$$G_{t+1} = I_t^g + (1 - \delta^g) G_t \tag{5.9}$$

where δ^g is the rate of depreciation of government capital. We simplify (5.9) by assuming that public capital fully depreciates over what we assume to be 20 year-long periods of the model. So, next period's public capital stock is determined solely by this period's public investment.

[2]For notational simplicity only, we assume the government's time discount factor is the same as that used by private households. One could allow the discount factor to differ from private households to study how the government's time preference affects policy.

To find the optimal fiscal policy, the government chooses sequences of tax rates, government consumption, and government capital to maximize the discounted utility of government officials and private households, given by (5.7), subject to a series of budget constraints and capital accumulation equations given above.[3] In addition, the government takes into account how their policy choices affect all private sector decisions. This only includes private capital accumulation, (5.6), since schooling and fertility, (5.1a) and (5.1b), are independent of fiscal policy. Finally, to obtain analytical solutions we assume $\delta = \delta^g = 1$, so that over our 20 year periods, both types of capital stocks fully depreciate. We also set $A = 1$, since its value is arbitrary in our calculations. The solution to the government's problem is (see the chapter Appendix for a sketch of the derivation)

$$\tau_t = \tau = \frac{(1-\alpha\beta)(1+\beta\mu\phi(1-\alpha)\Gamma)}{1+(1-\alpha\beta)\phi\Gamma}, \tag{5.10a}$$

$$g_{t+1} = \frac{\beta\mu(1-\alpha)}{(1+q)n_{t+1}} k_t^{\alpha} g_t^{\mu(1-\alpha)} \widehat{h}_t, \tag{5.10b}$$

$$k_{t+1} = \frac{\beta(1-\tau)(1-\alpha)}{(1+\beta+\psi)(1+q)n_{t+1}} \frac{k_t^{\alpha} g_t^{\mu(1-\alpha)} h_t}{\widehat{h}_{t+1}}, \tag{5.10c}$$

where

$$\Gamma \equiv 1+\beta+1+(\psi/\beta)+(\beta\alpha(1+\beta)+\beta(\alpha-1)+\psi\alpha)/(1-\alpha\beta).$$

Equation (5.10a) tells us the tax rate is constant over time. One can show that the constant tax rate τ is decreasing in ϕ, more concern for private households implies a lower tax rate. Equation (5.10b) gives a transition equation for the public capital stock that is analogous to that for the private capital stock. Here, the government's saving rate out of national income is a constant, $\beta\mu(1-\alpha)$. Combined with (5.10a) this tells us that a more selfish government, with a lower ϕ, will collect more in taxes but invest *a smaller fraction of tax revenue* in public capital—so as to maintain the *same* investment rate out of national income. Equation (5.10c) simply repeats the transition equation for private capital accumulation.

5.1.5 Steady State Equilibria

It is important to note that the model is recursive, which means that model can be solved in steps rather than all at once. The path for private sector schooling can be

[3] We assume that the government can commit to its policy choices in advance. For a discussion of commitment issues in regard to the setting of fiscal policy see Lundquist and Sargent (2004, Chapter 22).

determined first using (5.1b). Knowing the schooling path allows one to find a path for fertility using (5.1a). Finally, given the schooling-fertility dynamics, one can then determine the dynamics of government and private capital intensity using the difference equation system given by (5.10b) and (5.10c).

A country with sufficiently high initial schooling will experience growth and converge to a steady state as determined by (5.1) and (5.10). However, if $e_{t-1} = \bar{e}$, it may be the case that $\frac{\theta(\eta - \gamma T)}{\gamma(1-\theta)} < \bar{e}$. If this is true, then $e_t = \bar{e}$ and the economy is in a poverty trap where neither schooling nor fertility changes over time. For an economy with this initial condition, the only possible dynamics stems from the government and private physical capital accumulation in (5.10). Thus, different initial conditions may cause economies with identical structures to come to rest at very different steady state equilibria. One steady state has higher values of h, g and k, and lower levels of n, than the other.

It is also possible that economies differ in terms of the weight, ϕ, that their governments place on private household welfare in setting fiscal policy. Economies with higher ϕ will have lower tax rates, higher private capital-labor ratios, and higher levels of public capital. Higher ϕ causes higher worker productivity, even if the steady values of e and n are the same across economies.

In summary, worker productivity may differ either because of a poverty trap or because of policy differences. The next question is whether these sources of income differences are quantitatively important.

5.2 Cross-Country Income Differences

To investigate if the model is able to explain large income difference across countries, we compare the following two steady state equilibria, where e, n, g, and k are constant but different across countries. Our poor and rich steady states are characterized as follows.

Poor Country Steady State

(i) a binding poverty trap, $e_t = \bar{e}$ with high fertility, $n = 3.5$
(ii) a selfish government, $\phi^{poor} < \phi^{rich}$ with $\tau = 0.35$

Rich Country Steady State

(i) $e = T$ (full-time schooling) with low fertility $n = 1$
(ii) a low-tax government, $\phi^{poor} < \phi^{rich}$ with $\tau = 0.15$.

To quantify the model's predictions about income differences across these two equilibria, we calibrate the parameters to the rich country steady state. The physical capital income share, α, is set to the standard value of 1/3. The output elasticity for public capital, μ, is set to 0.30 . This value is somewhat less than the values estimated by Aschauer (1989) and Clarida (1993). However, the values of α and μ place the product $\mu(1-\alpha)$ at 0.2, an intermediate value of the estimates surveyed by Glomm and Ravikumar (1997). Based on Lord and Rangazas (2006), we set $\gamma = 0.28$ and $T = 0.50$. This implies potential earnings of a child that are about 14%of an adult's earnings. The annualized after-tax return to capital is set to 4%, the after-tax real rate of return to capital in the United States at the end of the twentieth century (Poterba (1997), Table 5.1)). The annualized rate of growth of exogenous technological change, q, is set to 1.0%. This is intended to reflect a worldwide, transferable, component of exogenous technological change.

The remaining parameters are set to match certain targets. We set ϕ^{rich} to match $\tau = 0.15$, about the ratio of government purchases to GDP in the United States. In the rich equilibrium steady state, we targeted $n = 1$, $e = 0.5$, and a value of k consistent with an annualized after-tax return of 4%. This means each household creates exactly one replacement household in the next generation, so the target is equivalent to assuming two children per couple in the data. The maximum schooling target can be interpreted as children going to school full time in the rich country.

In the poor country equilibrium, we targeted $n = 3.5$, which implies 7 children per couple. Despite the fertility decline in Africa since the 1980s, many of its poorest countries have Total Fertility Rates of 7 children per female (Bongaarts (2002)). In addition, the parameter settings must to be consistent with an optimal schooling level below $\bar{e}$. The minimal schooling level for young children is set to 0.08. This value implies that children in the rich country spend 6.25 times as much time in school over their childhoods than do children from poor countries. So if poor children spend 2 years in school, then rich children spend 12.5 years in school (assuming school years of equal length). Finally, as in Chap. 3, we set ϕ^{poor} in the poor country so that $\tau = 0.35$. Table 5.1 summarizes the parameter settings.

Table 5.2 presents the steady state worker productivity ratio, across rich and poor countries, generated by the model. The features included in the model cause the rich country to be over 28 times richer than the poor country. The table provides a decomposition of the worker productivity ratio based on the following expression for worker productivity,

$$y_t = \frac{k_t^{\alpha} A E_t g_t^{\mu(1-\alpha)} \widehat{h}_t}{1 + n_{t+1}(T - e_t)}. \tag{5.11}$$

Table 5.1 Calibrated parameter values

Parameter	Target
γ 0.2800	Relative child's earning
T 0.5000	Relative child's earnings
η 0.1646	Steady state fertility (poor)
θ 0.4049	Steady state schooling (rich)
α 0.3333	Standard value for capital share
μ 0.3000	Intermediate empirical estimate
ψ 0.2956	Steady state fertility (rich)
β 0.4999	Steady state return to capital (rich)

The numerator gives the various sources of production. The denominator gives the supply of labor per household, the adult worker and the child workers that are beyond school age.

The poverty trap causes the term $\widehat{h}_t/(1 + n_{t+1}(T - e_t))$, average human capital per worker, to be 3.7 times higher in the rich country for two reasons. First, since $e_t = 0.5$ in the rich-equilibrium and $e_t = \bar{e} = 0.08$ in the poor-equilibrium, adult human capital differs across countries. This causes output per worker in the rich country relative to that in the poor country to be 2.10, a value similar to that estimated by Hall and Jones (1999) using a much different approach. Second, the high fertility in the poor country implies that their workforce contains a sizeable fraction of young workers, who are less productive than adult workers due to less strength and experience (captured by $\gamma = 0.28$). This causes worker productivity to be 1.75 times higher in the rich country. The role of worker-age in determining low worker productivity is overlooked in most studies.

The poverty trap also causes low values of k and g. High population growth increases the size of next period's workforce relative to the current period's savers. High population growth spreads saving and capital accumulation more thinly across workers in the future, lowering k. Lower values of k and $\widehat{h}$ lower national income and reduce public investment.

The relatively low value of ϕ in the poor country raises tax rates and further reduces private saving and private capital formation. Indirectly this also lowers public capital formation by further reducing the level of national income. The combination of the poverty trap and the lower ϕ reduces public and private physical-capital intensities causing worker productivity to be 7.7 times higher in the rich country. This is over four times as high as the productivity ratio that Hall and Jones (1999) attribute to differences in capital intensity. There are several reasons why the estimate in Table 5.3 is higher.

In Table 5.2 we are assuming that the poor country is a perfectly closed economy. In the next section we open the economy to international capital flows. An open economy reduces the differences in capital intensity across rich and poor countries, although not completely. The typical poor country is neither perfectly open nor perfectly closed, so our estimates using perfectly closed and perfectly open economies should bound the estimate from Hall and Jones.

However, there are reasons to believe that the Hall and Jones estimates may be too low. Pritchett (2000) estimates that the actual capital stock in poor countries is between 57% and 75% of the officially measured stock. In poor countries the level of government consumption is under-estimated and the level of investment is over-estimated. This fact implies estimates of productivity differences that are based on direct estimates of capital stock differences, as in Hall and Jones, are too small.

The Hall and Jones approach also treats private and public investment as perfect substitutes in production. It is much more natural to assume that roads and private firms are complementary inputs. In addition, the estimates of the output-elasticity of public capital suggest the elasticity for public capital is about two thirds of the elasticity for private capital (Glomm and Ravikumar (1997)). Poor countries have

Table 5.2 Steady state worker productivity differential

Rich to poor ratios	Model prediction
k^{α}	3.68
$g^{\mu(1-\alpha)}$	2.09
$\widehat{h}/[1+n(T-e)]$	3.68
y	28.25

Table 5.3 Fiscal policy in the closed and open economy

Fiscal parameter	Closed economy	Open economy
τ	0.35	0.26
B	0.29	0.31
τB	0.10	0.08

relatively more public capital, implying that the perfect-substitutes assumption overstates the productivity of the capital stock in poor countries and lowers the estimated role of capital differences in explaining worker productivity differences.

5.2.1 Comments

The steady state comparison demonstrates that the sources of growth the model identifies are quantitatively important in explaining the large worker productivity gaps across countries. However, there are three reasons why the predicted gaps explained by these sources are overstated.

First, the predicted gaps are driven in part by the large difference in tax rates across rich and poor countries. While the countries in Table 3.1 of Chap. 3 have large governments, not all poor countries have this characteristic. For example, one could not use this reason for income gaps when looking back in US history. In the nineteenth century US tax rates were relatively low, so relatively high taxes and government consumption could not explain why the US was relatively poor in the nineteenth century compared to the end of the twentieth century. As in many poor countries today, explaining low worker productivity in the US past requires investigating other sources of poverty.

Second, the abstraction of our time periods causes the impact of fertility differences across countries to be too large. The problems stem from there being only one period of work for *both* adults *and* their children. As a result, the population of households that are currently saving compared to the future work force is unrealistically small because savers include only parents and future workers include only their children. Changes in fertility have a large effect on the relative sizes of these two groups. A high fertility rate, without a high rate of mortality, will imply a large increase in the size of the future workforce, causing the capital accumulation financed by the current period's saving to be more thinly spread over the next generation of workers than in a model with many periods of work (and saving) and with the high death rates that mediate population growth in poor countries. In the real world parents and adult children *both* work and save for many periods, along with many other unrelated age-cohorts of workers and savers.

Third, as mentioned, we have assumed that the poor country is perfectly closed. If we instead assume a perfectly open economy, as we will do later in the chapter, the income gaps will not be nearly as large because physical capital will flow into the poor country from abroad. Most poor countries are imperfectly open, so the two extreme assumptions will bound the actual situation.

While the sources of poverty we have identified are important, they likely do not explain the full extent of the poverty in many developing countries. This motivates the second half of the book, where we use a two-sector approach to identify additional sources of poverty. The two sector approach will provide new reasons why fertility is high and saving is low in poor economies. It also identifies a common source of labor market inefficiency in developing countries that keeps average worker productivity low.

5.3 International Financial Institutions and Foreign Aid

International economic assistance is a prominent feature of the global system since the 1950s. National governments in advanced countries and international organizations jointly owned by governments provide loans and grants to developing nations and other countries in need due to mismanagement, conflict, natural disasters and other bad luck.

The most prominent of the international financial institutions (IFIs) are the International Monetary Fund (IMF) and the International Bank for Reconstruction and Development (IBRD), which is commonly known as the World Bank. These two US-based financial institutions (IFIs) were created in the 1940s. They have near universal membership of 188 member countries. The IFIs' broad purpose, and of what eventually became the World Trade Organization, is to help underpin a peaceful, market-oriented international economic system and avoid, as the IMF's charter puts it, "policies that are destructive of national and international prosperity". Inefficient policies are sometimes optimal from a narrow national perspective, but are zero- or negative-sum games from an international perspective. In the 1930s, for example, tariff wars, competitive devaluations, and other policy mistakes was a natural response of countries trying to shield their employment from foreign competition. But, collectively, these policies resulted in persistently high global unemployment and economic stagnation in the 1930s. This contributed significantly to political instability in major countries, including Germany, and thus helped trigger World War II.

The broad division of labor in the IFIs is as follows: the World Bank supports long term development through program and project loans and grants while the IMF focuses on macroeconomics and macro-critical structural reforms in its member countries. IFI programs are complemented by those of regional development banks (African Development, Asian Development Bank, etc.), as well as bilateral aid provided by national governments directly to countries in need. In the United States, for example, food aid is administered by the US Agency for International Development. Canada and all major advanced countries have ministries or agencies devoted to international development assistance.

International assistance does not stop with the transfer of financial and real resources but is complemented by various specialized technical services. It is recognized that to overcome poverty and underdevelopment requires more than money and resources. For example, countries may not collect enough revenue to be able to pay for badly needed government services, including basic sanitation, health and education. International technical assistance aims to help countries build their institutions—fix broken tax systems, set up central banks and manage sound money, and otherwise help nurture economic institutions that can help countries run more efficiently.

5.3.1 Conditionality and Ownership

With the exception of humanitarian aid, international assistance often comes with strings attached. *Conditions* aim to help recipient countries improve their policy choices and economize on the donors' or creditors' resources so that the money can be spread to more needy countries and worthy projects. International loans are sometimes backed by collateral, but this is not typically the case for IMF or World Bank loans. Lacking collateral, conditions help assure the international community that these loans will be repaid. Conditions also help give borrowing countries reasonable certainty. They are assured that they will continue receiving international assistance if they meet conditions specified (this is called uniformity of treatment).

The IMF got involved in structural conditionality in the 1980s for good reasons. Supply side reforms improve the efficiency of the economy and boost exports and growth. They also prevent situations where governments cut the wrong components of spending or follow other inefficient policies. But while well intentioned, the expansion in the scope and complexity of Fund conditionality had unintended consequences. The Fund was spread thin. The content of its programs expanded far beyond macroeconomic and financial stabilization – its traditional areas of expertise. Involvement in complex, multi-stage structural reforms led the IMF away from its core areas in which it possesses a clear comparative advantage and made it lose focus.

Domestic ownership for structural reforms was undermined. Cash-strapped governments sometimes agreed to IMF conditionality mainly to access IMF financing and obtain its seal of approval. Lack of ownership hindered policy implementation and undermined the IMF's credibility and its ability to catalyze reforms. Some politicians learned to treat the Fund as a scapegoat for the tough choices they had to make. Conditionality came to be viewed as an inevitable but much-resented sacrifice of national sovereignty rather than an instrument of international cooperation. Such rhetoric helped galvanize opposition to needed reforms, especially when they were perceived as being imposed from abroad to protect the narrow interests of donors. It became clear that unless the country “owned” the policy reforms, i.e. believed in and supported the implementation of the reforms, there was little chance of success.

The Fund’s deepening involvement in structural areas also caused frictions with the World Bank. The roles of the two Bretton Woods institutions converged during the 1980s. The Bank’s interest in macroeconomic developments grew after it added structural adjustment lending to its operations following the debt crisis. As the Fund started worrying about structural issues much more in the late 1980s and 1990s, the overlap in the activities of the two institutions increased. Coordinating the activities of the two institutions became a higher priority.

As the Fund and the Bank appreciated the importance of strong ownership for the implementation of structural reforms and for market confidence, they reassessed their approach in low-income countries. In 1999, a transparent, country-driven process was introduced involving tripartite collaboration between the country, the Fund and the Bank. The country draws up its Poverty Reduction Strategy Paper (PRSP), laying out its programs. This is supposed to be done in consultation with the country’s own population, the multilaterals and the bilateral donors. This is a partnership between the various groups, but primarily it is the country itself that’s expected to draw up the program and take ownership of it.

The IFIs have made important changes in their conditionality in recent years. Following input from policy makers, civil society and academia, the IMF now aims to have conditions meet five principles: (i) national ownership of programs, (ii) parsimony in conditionality; (iii) tailoring of programs to borrowing country circumstances; (iv) coordination with other multilateral institutions; and (v) clarity in conditionality. Importantly, IFI inclusive processes of transparency, consultation, and persuasion aim to empower members to design and implement programs and make their dialogue with the Fund authentic. IMF staff members are encouraged to seek at an early stage in negotiations proposals from country authorities and to be flexible in program design. The timing of programs is to become more flexible in order to deal with situations in which time is short or the authorities’ capacity is limited.

The Fund aims to help countries build broad support for sound policies, including through public discussion and the adoption of participatory processes (see also Drazen and Isard (2004)). IMF staff is encouraged to assist through seminars, meetings with parliamentary committees, trade unions, business groups, and the media. IMF resident representatives are to play a key role in this.

5.3.2 Summary

Left on their own, some countries’ economic and political systems generate short-sighted and inefficient policies that harm their national welfare and have systemic effects—meaning regional and global spillovers. The IFIs were created in the 1940s to help countries in need. Their financing and technical support allows countries to discover and adopt policies that are more efficient, more sustainable, and have fewer negative spillovers. Conditions attached to international assistance are simply the way for the international community to ensure that countries it shows solidarity to are directing their efforts in the right direction.

Low country ownership of IFI-supported programs in the 1990s was the result of overreaching of conditionality. Analytical work helped assess the problem by focusing on the incentives of policy makers undertaking reforms, the resistance of interest groups to reforms, and on difficulties in monitoring and implementing reforms. Conditionality must be tailored and targeted to take into account the domestic political economy so as to maximize ownership. In practice, this means taking into account the influence of special interest groups, other domestic divisions, and the state of institutions in countries being assisted. The reevaluation of conditionality triggered by the difficulties with programs in the 1990s led to important reforms. The focus of conditionality has been sharpened, and the aim is to use it in absolutely critical situations—to tackle first order problems, not as an opportunity to pass desired but otherwise noncritical reforms.

5.4 Foreign Aid and Policy Experiments

Section 5.2 identified some potentially important sources of income differences across countries. Section 5.3 broadly discusses the attempt to use international aid to increase growth in poor developing countries. Evidence suggests that these attempts have not been generally successful. In this section, we search for policies that can effectively eliminate the sources of poverty and generate growth that would allow the living standards in the poor country to converge to those in the rich country. We begin with a standard form of unconditional aid as our baseline for comparison—government budget support. We then consider four policies: opening the economy to international capital flows, two education policies aimed at eliminating the poverty trap, and a policy reform designed to eliminate the poor country's anti-growth domestic fiscal policy. While there are other policies that might be considered, our model is best suited to analyze these policies. The results of our analysis will reveal many of the issues associated with policy reform and aid in general.

5.4.1 Unconditional Aid—Budget Support

We first analyze *unconditional* aid that takes the form of budget support to the poor country's government. Budget support will serve as a baseline to compare against other aid policies that are *conditional* in the sense that they are tied to specific policy changes.

Radelet et al. (2006) report that current aid-flows average about 5% of the recipient countries' GDP. Our model is calibrated to match poor countries with large governments where government purchases comprise about one third of GDP. For these countries, the average aid flow is 15% of the net tax revenue used for

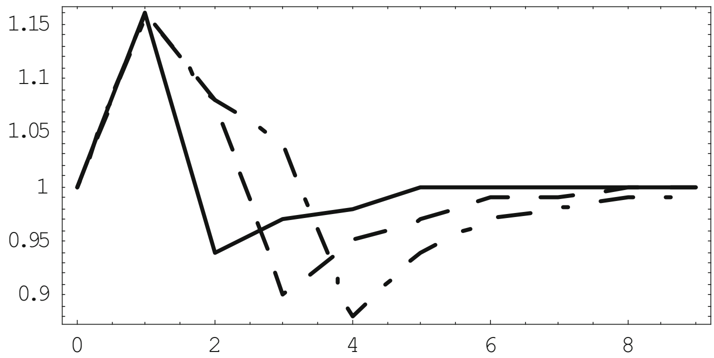


Fig. 5.1 Worker productivity with unconditional aid
Notes: Fig. 5.1 shows annualized growth rates in worker productivity over time for unconditional aid policies beginning in period 0. The aid provided is 15% of the government budget. The solid line represents the effects of giving aid for a single period; the dashed line shows the effects of aid provided for two periods, while the dash-dot shows the effects of aid provided for three periods.

government purchases. We consider aid flows equal to 15% of net tax revenues with varying duration: one, two, and three periods (or 20, 40, and 60 years). The impact of these aid flows on the growth rate in worker productivity is presented in Fig. 5.1.

The initial steady state growth rate of the economy is 1%, the exogenous annualized rate of growth due to technological progress. The aid inflows increase growth rates initially, but only by modest amounts. In the initial period, annualized growth rates rise to 1.15%. The modest initial increase in growth rates results from the fact that the government will save and invest a fraction of the aid causing public capital to increase. Greater public capital raises the marginal product of private inputs and the rental rate on human capital, which raises private saving and private physical capital accumulation.

After the first period, growth rates fall. The economy is unable to sustain even the modest increase in growth rates for two reasons. First, since the aid flow is only temporary, the rise in public saving cannot be sustained. Second, there are diminishing returns to public and private investment that would cause growth rates to decline back to the steady state level, even if aid inflows were permanent. Growth rates eventually dip below the steady state level for several periods because the rise in the public and private capital intensity cannot be sustained and the economy must revert back to the initial steady state capital intensities. In short, unconditional aid temporarily, but not permanently, shifts the economy's transition equations upward. With no permanent structural change in the economy's dynamics, it must return to its original steady state. The empirical analysis of Radelet et al. (2006) shows budget support raises growth rates temporarily. However, our model suggests that there are no long-run income benefits from unconditional budget support.

5.4.2 Opening the Economy

Up until now we have assumed that the poor country's economy is perfectly closed. As we saw in Sect 5.3, IFIs encourage countries to open their borders. What happens if the poor economy is opened to trade and international capital flows? What will be the effect on different generations of households in the poor country? Will opening the economy make the poor country's government better off or will it oppose the policy?

To answer these questions, the model must first be re-solved under the assumption that the economy is open. In an open economy, private capital flows will equate the poor country's interest rate to the exogenous world interest rate (which we take to be the steady state interest rate in the rich country). Next, the dynamic path under the open economy assumption is computed as the economy goes through the transition from the initial closed economy steady state to the open economy steady state. Unlike most neoclassical growth models, the "small" poor economy will *not* "jump" to the new steady state in a single period as interest rates are equalized in an open world capital market. This is because, unlike private capital, the government's *public* capital accumulation will adjust *gradually* to the opening of the economy. Public capital (such as roads and public utilities) does not flow across borders in the same way that private capital does. Note from Eq. (5.4a) that interest rates can be equalized due to private capital adjustments alone. Finally, welfare comparisons are made to see who benefits and who loses from opening the economy, an analysis that includes computing the welfare effect on the poor country's government itself.

After the economy is opened, the poor country's after-tax rental rate will converge to the world after-tax rental, $r*$, which we take to be the steady state interest rate of the rich country. The equilibrating force is assumed to be private capital mobility. The poor country's private capital intensity is determined by using (5.4a) and the international capital market condition $r* = (1 - \tau_t)r_t$. Note that this does not mean that k is equated across rich and poor countries because g may differ across countries. Smaller values of g lower the marginal product of k, so smaller values of k are needed to drive the return to physical capital down to the world interest rate.

With k determined internationally, the government's optimal fiscal policy will also change. The government now maximizes (5.7) not subject to (5.6), as in the closed economy, but subject to the k determined by international capital markets as described above. The optimal policy in an open economy becomes (see the chapter Appendix for the derivation)

$$\tau = \frac{\beta(1-\alpha)}{\beta + \phi[\psi + \beta(1+\beta)]} \tag{5.12a}$$

$$g_{t+1} = \frac{B\tau}{(1+\mathrm{q})n_{t+1}} \left(\frac{\alpha(1-\tau)}{r*} \right)^{\frac{\alpha}{1-\alpha}} g_t^{\mu} \widehat{h}_t \tag{5.12b}$$

where

$$B = \frac{\mu\beta + \phi\mu(\beta(1+\beta)+\psi)}{1+\phi\mu(\beta(1+\beta)+\psi)}.$$

The coefficient B in the transition Eq. (5.12b) represents the share of the government budget that is invested in public capital. The product $B\tau$ is the share of national output that is invested in public capital.

We now compare the poor country's fiscal policy in open and closed economies. Begin by considering the extreme case where $\phi = 0$. Comparing (5.12) to (5.10), we note

$$B^{open} \equiv \beta\mu > \beta\mu \frac{1-\alpha}{1-\alpha\beta} \equiv B^{closed},$$

$$\tau^{open} \equiv 1-\alpha < 1-\alpha\beta \equiv \tau^{closed},$$

$$(B\tau)^{open} = \beta\mu(1-\alpha) = (B\tau)^{closed},$$

where all inequalities hold provided that future utility is discounted, i.e. provided that $\beta < 1$.Opening the economy raises the portion of the budget that is invested and lowers the tax rate, but leaves the fraction of national output invested the same.

The fiscal policy differences are due to the *timing* of the impact of fiscal policy on private capital formation in open versus closed economies. In a closed economy, government policy affects private capital formation by affecting the after-tax wage of savers that fund the *next* period's private capital intensity. In an open economy, government policy affects private capital intensity by affecting the marginal product of private investments in the poor country—reducing it with higher tax rates and raising it with higher public capital intensity. International capital flows will anticipate and respond to these changes in private returns to investment, until the after-tax return to investment are equalized across countries. Thus, in an open economy, government policy has a more *immediate* effect on private capital formation—this period's policy affects *this* period's capital intensity rather than this period's saving flow and *next* period's capital intensity (as in a closed economy). With discounting of the future ($\beta < 1$), the cost of high taxes and low public investment, in lowering private capital intensity, is smaller in the closed economy due to the one-period delay in their effect. In this sense, opening the economy makes private capital formation more responsive to policy changes. The government reacts to the new environment by choosing a more "pro-growth" fiscal policy stance.

Table 5.3 gives the fiscal policies in open and closed economies for the calibration in Table 5.2 where, instead of $\phi = 0$, we have $\phi = 0.7461$. The result with $\phi = 0$ extends to positive values of ϕ; taxes are lower and the fraction of the government budget invested is higher in an open economy. However, the share of national output that is invested in public capital is lower in the open economy when $\phi > 0$, as B rises less than τ falls when the economy is opened. Thus, opening the economy lowers the economy's rate of public investment out of national output.

Figure 5.2 shows the effects on worker productivity of opening the economy to foreign investment. Growth accelerates in the first period as the capital inflow narrows gap in private capital intensity across rich and poor countries. The capital inflow raises the recipient country's national income and tax base, offsetting the reduction in the rate of investment in public capital. Public capital intensity rises over time to a higher steady state value. The increase in public capital intensity raises the marginal product of private capital and causes private-capital intensity to increase further. The modest additional increases in public and private capital intensities keeps growth in worker productivity above the rate of technological change until period 4, when the economy has approximately converged to its new physical capital intensities.

The growth effects of opening the economy to capital mobility dwarf those of the unconditional aid policy. Moreover, these effects are permanent in nature because the change in the economy is structural. The new steady state is characterized by higher permanent per capita incomes.

The extent to which inflows of private capital narrow productivity differences in the long-run is given in Table 5.4; the counterpart to Table 5.2 in a perfectly open economy. Comparing Table 5.4 to Table 5.2, one sees that worker productivity gaps are narrowed by opening the economy. The rich country's advantage in worker

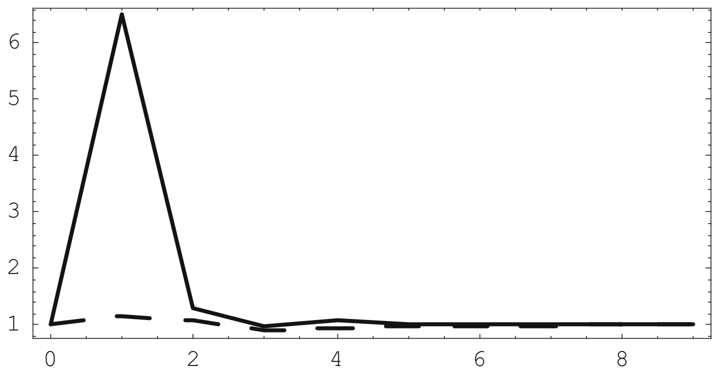


Fig. 5.2 Opening the economy to capital flows
Notes: Figure 5.3 plots give the annualized growth rates in worker productivity over time from opening the economy compared to a two-period unconditional aid policy. Unconditional aid flows are 15% of government budgets in each of the two periods. Solid line—open economy, Dashed line—two periods of unconditional aid.

Table 5.4 Steady state productivity differences—open economy

Rich to poor ratios	Model prediction
k^{α}	1.41
$g^{\mu(1-\alpha)}$	1.73
$\widehat{h}/[1+n(T-e)]$	3.68
y	8.98

productivity is now less than 1/3 of what it was in a close economy setting, although a nine fold difference still remains.

While there are clear gains in worker productivity from opening the economy, not all generations benefit from the opening. The policy affects the welfare of households by affecting factor prices. Households prefer higher current wages for themselves and higher future wages for their children. They also benefit from higher interest rates on their life-cycle saving. Opening the economy will raise wages and lower interest rates as capital flows into the economy. For *most* generations there is a net gain in utility from these factor price adjustments (the effect of higher wages is greater than the effect of lower interest rates). This is *not* true for the initial generation of young households who are alive at the time the policy is introduced. Their current wages are unaffected by the capital inflows (since the initial capital intensity is fixed) and yet their interest rates are significantly lowered. The sharp drop in interest rates, with no change in current wages, causes their welfare to fall. Thus, welfare falls for the first generation and rises for all others.

The government in the poor country enjoys an increase in public consumption each period—the increase in the tax base from capital inflows offsets the drop in tax rates. The gain in the government consumption, along with the discounted gain in utility to all future generations, is larger than the loss in welfare of the initial generation. Thus, the poor government would want to open the economy, on economic grounds, in our setting.

This finding is obviously sensitive to the particular calibration chosen. If the initial capital intensities were smaller, or if the poor country's government had a higher rate of time preference, then one might see opposition to opening the economy.

5.4.3 Eliminating the Poverty Trap

Schooling is low in the poor country because the value of forgone earnings associated with sending older children to school is a relatively large fraction of parent's income. The value of forgone earnings is high because households have many children and because parental earnings are low. The poverty trap can be removed if parental earnings are increased relative to the earnings of older children. This would make it more costly to have many children (because of the lost work time and forgone consumption of parents associated with child rearing) and it would lower the relative value of children's work in total family income.

Using aid to encourage poor countries to increase the schooling of *younger* children (i.e. to increase $\bar{e}$) will increase earnings but will not remove the poverty trap. This type of policy does not raise the earnings of parents *relative* to those of older children (since they both receive the higher levels of education when they are young children). What is needed is more schooling of *older* children, so that when they become parents their earnings (based on $e_t > \bar{e}$) exceed the earnings of their older children (based only on $\bar{e}$)—thereby making children more costly and relatively less important in generating family income.

One policy that can remove the poverty trap is similar to Mexico's Progresa program.[4] Under this policy, the governments subsidizes the forgone earnings of older children who attend school. A sufficiently high subsidy would raise e_t sufficiently above $\bar{e}$, so that transitional dynamics would result, sending the poor country to the high schooling steady state. A potential advantage of identifying and eliminating poverty traps is that aid need not be ongoing. Once sufficient aid has been provided to eliminate the poverty trap, no further aid is necessary.

To begin the analysis of the subsidy policy, introduce the policy parameter ν that indicates the fraction of forgone earnings of older children that the government returns to the household. This introduces the expression $\nu w_t \gamma \bar{h}_t (e_t - \bar{e}) n_{t+1}$ on the right-hand side of the household lifetime budget constraint from Sect. 5.1. In presence of the subsidy, household behavior becomes (see *Problem 7*)

$$n_{t+1} = \frac{\psi}{(1+\beta+\psi)\left(\eta - \gamma\left(T - e_t + \nu\left(e_t - \bar{e}\right)\right)\left(\bar{e}/e_{t-1}\right)^{\theta}\right)} \tag{5.13a}$$

$$e_t = \max\left[\frac{\theta\left(\eta\left(e_{t-1}/\bar{e}\right)^{\theta} - \gamma T + \gamma \bar{e}\nu\right)}{\gamma(1-\theta)(1-\nu)}, \bar{e}\right]. \tag{5.13b}$$

The subsidy increases the optimal schooling level and, if the subsidy is sufficiently high, the optimal schooling level is pushed above $\bar{e}$. For a given level of e_t, fertility is also encouraged by the subsidy. A rise in fertility has the unintended consequence of lowering growth. However, if the subsidy raises e_t enough, then fertility will fall.[5]

[4] In 1997 Mexico began *Progresa*, a program designed to increase human capital in poor families by paying families to send their children to school and to visit health care providers. Grants are provided directly by the government to the mothers of children. The school grants cover about 2/3 of what the child would receive in full time work (Krueger 2002).

[5] Our model abstracts from tuition costs (see *Problems* 2 and 3 from Chap. 4). The government can raise schooling by increasing tuition subsidies. Doepke (2004) and Lord and Rangazas (2006) study the historical impact of government tuition subsidies in England. They find that lower tuition has modest effects on schooling and growth. Lower tuition reduces the cost of all children and, in particular, young children who would have attended school in any case. This raises fertility for several periods and slows the demographic transition. Thus, something like a Progresa program or compulsory schooling is needed to generate a quick demographic transition and rapid economic growth.

Of course, the subsidy must be financed out of tax revenues. In addition, as older children work less in order to attend school, the tax base shrinks. So, government revenue is reduced by two factors in the first period—the subsidy payment and the decline in the tax base. This implies that government consumption and investment will fall initially. As with the possible rise in fertility, reduced government investment may offset the early growth effects of the policy. As the stock of human capital rises and increases the tax base, government consumption and investment will eventually rise.

Table 5.5 and Fig. 5.3 present the effects of the Progresa program with a subsidy, lasting for a single period, that is similar in size to that offered in the Mexican Progresa program ($\nu = 0.67$). The relatively large one-period subsidy is more than enough to boost the economy out of the poverty trap and in fact creates something close to a "growth miracle." The large rise in schooling is enough to create a fall in fertility. Rising human capital per worker also increases physical and public capital intensities generating growth for a number of periods. As in the case of opening the economy, the growth effects dwarf those of budget support and lead to large permanent increases in income levels.

Table 5.5 The progresa program: schooling and fertility effects

Period	1	2	3	4	5	6	7	8	9	10
e	0.29	0.34	0.37	0.41	0.43	0.45	0.47	0.48	0.49	0.49
n	2.5	1.2	1.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0

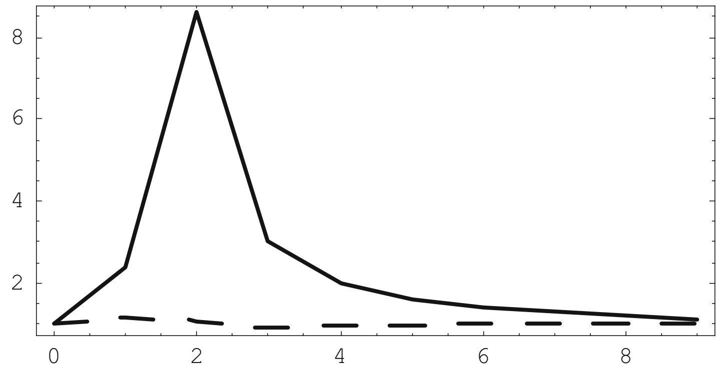


Fig. 5.3 The Effects of a progresa program
Notes: Fig. 5.4 plots give the annualized growth rates in worker productivity over time from the Progresa program compared to a two-period unconditional aid policy. Unconditional aid flows are 15% of government budgets in each of the two periods. Solid line—Progresa, Dashed line—two periods of aid.

An advantage of the Progresa program is that no generation is hurt by the policy. Although the positive welfare gain is quite small for the first generation, since parents miss the direct benefit of higher schooling, it is significant from the second generation on. With sizeable welfare gains after the first period, combined with large increases in the tax base, the government's welfare increases due to the policy change.

Other policies designed to increase schooling do not necessarily generate universal private sector benefits. We compare the welfare effects of the Progresa policy to a compulsory schooling policy, frequently seen in practice, that generates the same increase in schooling. This requires that the first generation of students spend 0.29 of their time endowment in school, as under the Progresa policy. After that point, the minimum school requirement of 0.29 is nonbinding and schooling will follow the same path that is displayed in Table 5.5.

The growth effects of compulsory schooling are actually stronger than the Progresa policy. This is because, without the government subsidy, families will choose fewer children relative to the Progresa policy that subsidizes the cost of schooling children. The steeper decline in fertility increases the economy's growth rates marginally above those in Fig. 5.5. However, since the initial family is forced to send their children to school more than what they find optimal, they are made worse off. The government, on the other hand, prefers compulsory schooling. The fact that compulsion eliminates the need for a subsidy and raises growth and tax revenue to a greater degree, more than compensates for decrease in welfare for the initial generation.

5.4.4 Fiscal Policy Reform

Attempting to reform conventional fiscal policy of developing countries is a common target for aid policy. We now consider the effects of imposing a fiscal policy in the poor country that would bring it in line with the fiscal policy of the rich country. In particular we compute the effects of imposing the $\tau$and B of the rich country, where the optimal values are 0.15 and 0.67, on the poor country, where the corresponding optimal values in the open economy are 0.26 and 0.31.

The effect of fiscal policy reform on the growth rates of worker productivity are given in Fig. 5.4. The growth effects are relatively modest and short-lived. In part, this is due to the fact that we begin the policy experiment from a perfectly open economy. Opening the economy brings the fiscal policy of the poor government closer to that of the rich government (see Table 5.4). This has the effect of making the differences in tax policy less dramatic and the returns to accumulating private and public capital smaller (since poor-country capital intensities are higher in the open economy than in the closed economy). When the poor economy is relatively close to the rich country's capital intensities to begin with (see Table 5.4), the transition to new steady state is short.

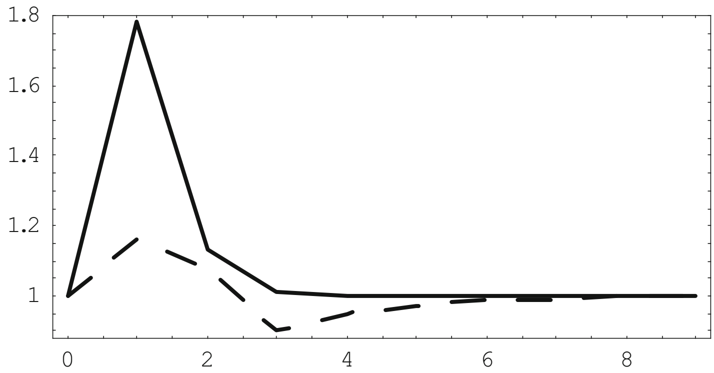


Fig. 5.4 The Effects of fiscal reform
Notes: Fig. 5.5 plots give the annualized growth rates in worker productivity over time from the fiscal reform compared to a two-period unconditional aid policy. Unconditional aid flows are 15% of government budgets in each of the two periods. Solid line—fiscal reform, Dashed line—two periods of aid.

There is, however, a significant gain in utility of all generations from the fiscal reform. This is because of the growth effects highlighted in Fig. 5.4 and because of the direct benefits of paying lower taxes. Of course, the welfare of the poor country's government falls significantly since they have been moved off their optimal fiscal policy.

5.5 The Aid Cost of Reform

We have examined five policies that might be used to promote growth in developing economies. The impact of the policies on growth differed significantly and so do their aid cost. The unconditional aid policy comes at a price and delivers no long-term benefits. Openness and the Progresa-style education subsidy deliver large and sustained increases in income. They also increase the welfare of the poor country's government and thus should be readily accepted. However, openness hurts the initial generation of private households, and thus may not increase the government's welfare for all calibrations. At a minimum, the government may use the fact that the current generation is hurt as a "bargaining chip" to induce some aid compensation for opening the economy. Strategic considerations also enter in the case of the Progresa program. The government prefers compulsory schooling and they may use this as a threat point to induce aid compensation for going forth with the Progresa program.

The domestic fiscal reforms, unlike the other policies, would certainly be opposed by the poor country's government. Aid dollars would have to be used to "purchase" the fiscal reforms from the poor country's government, in compensation for its

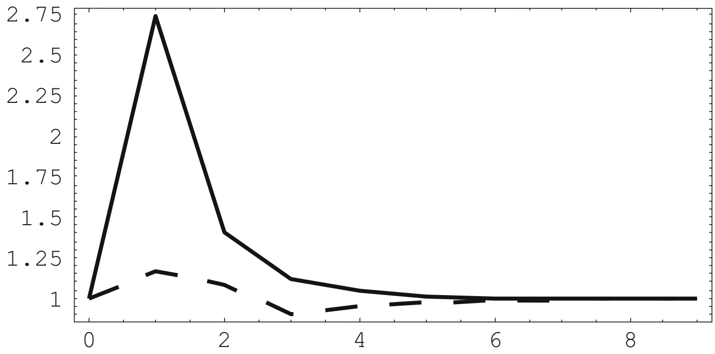


Fig. 5.5 Fiscal reform with required aid
Notes: Plots give annualized growth rates in worker productivity over time from the fiscal reform versus a two-period unconditional aid policy. The required aid is a permanent flow equal to 87% of the government budget. Unconditional aid flows are 15% of government budgets over each of the two periods. Solid line—fiscal reform, Dashed line—two periods of aid.

losses. We can assess the aid cost of fiscal reform by calculating the minimum amount of aid needed to keep the poor country's government indifferent to the reform. We compute the aid cost as a permanent flow of aid, expressed as a fraction of the poor government's budget. The aid flow must be permanent because the government will want to renege and revert back to its optimal fiscal policy as long as it stays in power. Of course, the aid flow will also change the amount that the government invests (while the government consumes most of the aid flow, some is invested) and thus the growth effects of the fiscal reforms will be larger than those without aid—an added benefit of the aid that goes beyond purchasing the reforms per se. The growth effects are given in Fig. 5.5.

As mentioned, the growth effects are higher than in Fig. 5.4 because the government chooses to invest some of the aid. However, the amount of aid required to purchase the reform is very high. Aid equal to over 87% of the poor country's budget is needed. Since the poor country's budget increases as the country grows, the absolute flow of aid must increase over time—long after the growth rate effects of the reforms have been exhausted.

5.6 Aid Failures

It is a discouraging stylized fact about development that no robust correlation between aid and growth has been identified in the econometric literature (e.g. Easterly et al. (2004)). There are several possible econometric reasons for the absence of a clear positive relationship. For example, *endogeneity of aid flows* (aid is targeted to slow-growing economies), *specification error* (the relationship between aid and growth is highly nonlinear) and *measurement error* (all aid, including aid not intended to generate growth, is lumped together in a single measure). Attempts to account for these econometric issues have failed to overturn the consensus finding of a zero correlation between aid and growth (see, for example, Rajan and Subramanian (2008, 2011). Our analysis is consistent with three possible reasons for the lack of correlation.

5.6.1 Unconditional Aid Is Not Growth-Promoting

Our results suggest that unconditional aid, including aid where conditions are not adequately enforced, will not deliver long-term gains in income. The boost to growth from unconditional aid is short-lived and so modest that it could easily be overshadowed by other developments – e.g., any long-lasting effects of the negative shocks to the economy that initially triggered the scaling up of unconditional aid to begin with.

5.6.2 Domestic Conflict Over Growth Policies

While there are policies that can generate rapid growth and sustained increases in income, there is likely to be domestic conflict over which policy to pursue. The government favors opening the economy and compulsory schooling, but the current generation of private households will oppose both policies. The current generation of private households favors the Progresa program, a program which the government views as clearly inferior to compulsory schooling. These conflicts may undermine attempts to achieve domestic consensus over which growth-promoting and poverty-reducing policies to implement. Such lack of consensus could delay or undermine the negotiation and implementation of conditional aid agreements with donors.

5.6.3 Prohibitive Aid Cost

Fiscal reforms are often part of the conditions for receiving aid. Our analysis suggests that reforms of domestic fiscal policy are likely to be the least successful of the policies that we examined. First, the growth effects of fiscal reform are relatively modest and short-lived. Second, the aid-cost of "buying" the reforms from the poor country's government are enormous. Unless the aid keeps flowing to the poor country in sufficient quantity, the domestic government will do what it can to revert back to a high-tax, low-investment regime. In fact, the cost of maintaining effective reforms will increase over time as the government's budget, and the potential to increase government consumption, grows. In practice, aid is likely be far less than what is necessary to keep the government indifferent and thus fiscal reforms may be doomed from the beginning.

Even if the aid is carried out in sufficient amounts indefinitely, there will be little correlation between aid and economic growth in the data. The growth effects occur early on, while the aid continues into the future during periods where the growth effects have long since vanished. If aid is reduced or maintained at constant level, rather than increased, there will be a reversion in fiscal policy and growth. Thus, aid could be flowing to a country experiencing negative growth.

5.7 Humanitarian Aid

We have seen that there can be significant difficulties associated with using foreign aid to finance large-scale government investment projects or to induce changes in national policy. When the confidence in the foreign government is low, it is natural to think of bypassing the government in the seemingly less ambitious goal of providing humanitarian aid directly to the country's poor. Unfortunately, even providing food to the poor people of a country is difficult.

In a recent paper, Nunn and Qian (2014) summarize the problems with providing humanitarian aid. Food aid is often diverted to the government or stolen in route by rebel or criminal groups within the country. The problem becomes particularly acute when there is ongoing unrest and conflict in the country. In this type of setting humanitarian aid is accused of not only being ineffective, but also of promoting conflict.

Nunn and Qian do a careful econometric analysis of the issue. They find that food aid increases the *duration* of civil wars in countries that have a history of internal conflict. They find no evidence that food aid *starts* civil wars or wars with neighboring countries. This suggests that, during civil wars, the combatants are able to divert the food to soldiers and away from the population in general. Thus, just as with other types of aid, the country-specific situation has to be right for aid to have a chance of working in the way the donors intend.

5.8 Conclusion

This chapter introduces three extensions to the standard one-sector neoclassical growth model of physical capital accumulation: endogenous theories of human capital accumulation, fertility, and fiscal policy formation. These extensions are needed to provide a more complete theory of both historical growth within a country and the large income differences across countries. The role of private saving, emphasized in the standard model, is certainly important. However, the majority of economic growth cannot be explained without also considering the effects of investment in schooling and public infrastructure, private sector taxation and government consumption, and population demographics associated with fertility choices.

The extended growth model also allows us to think about the fundamental sources of poverty traps and anti-growth fiscal policy. Identifying these important fundamentals of low income is needed to assess the likely impact of various policy suggestions intended to trigger growth. We saw that the effectiveness of different policies varies dramatically. Political support for different policies from within the recipient country was also shown to vary widely. Even in our simple model, there are clear economic and political reasons why some of the more popular policy recommendations typically made are not likely to generate sustained growth. Difficulties in finding policies that both (i) fundamentally alter the country's long-run growth potential and (ii) receive strong support from the recipient country's government and powerful interest groups explain the disappointing effect of foreign aid on growth in developing countries. In practice, working with recipient countries to identify pro-growth policies with strong political support will make foreign aid more effective.

5.9 Related Literature

Much of the chapter is based on Mourmouras and Rangazas (2007). Several books on economic growth have recently been written with a focus on human capital and fertility. An introduction to human capital and fertility that is appropriate for undergraduates can be found in Lord (2002) and Weil (2009). Two excellent advanced books on the subject are Galor (2011) and de la Croix (2013). The theory in these books is aimed at graduate students and research economists, but there is much material that any serious students of economics could benefit from.

For more on the growth consequences of missing intergenerational markets for human capital loans, including in some cases the connection to poverty traps, see Drazen (1978), Azariadis and Drazen (1990), Becker et al. (1990), Galor and Ziera (1993), Rangazas (2000), Acemoglu (2009, section 21.6), and Cordoba and Ripoll (2013). Azariadis and Stachurski (2004) provide an extensive survey and assessment of poverty traps.

The political economy of policy formation is a growing field in economic theory. A nice non-technical introduction to the political economy of growth is Acemoglu and Robinson (2012). Galor (2011) and de la Croix (2013) contain political theories of public education based on economic considerations. Ivanyna et al. (2018) offer an introduction to political economy issues associated with economic growth. More advanced texts that provide surveys of political economy more generally include Persson and Tabellini (1990) and Drazen (2000). Grossman and Helpman (2001) focus on the economics of interest group politics in particular.

Popular introductions to the issues associated with foreign aid and growth include Easterly (2001, 2007), Mallaby (2004), and, the slightly more technical, Isard et al. (2006). Most academic work on aid is empirical, but there are advanced journal articles that address various theoretical concerns such as Svensson (2000), Mourmouras and Mayer (2005), Marchesi and Sabani (2007), and Scholl (2009).

5.10 Exercises

Questions

1. Describe the complete one sector growth model. What are the key determinants of growth?
2. How do the private stocks of physical capital affect the accumulation of public capital?
3. What does it mean to say that a model is recursive? In what sense is the model recursive?
4. How does a more selfish government undermine private capital accumulation?
5. What are the fundamental differences in the economies of the “rich’ and the “poor” countries?
6. How are ϕ^{rich} and ϕ^{poor} calibrated?
7. Summarize the results of Table 5.2 where the worker productivity gap between rich and poor countries is estimated and decomposed into its contributing factors.
8. Compare the model’s explanation for productivity gaps between rich and poor countries to the explanation of Hall and Jones (1999).
9. Discuss the ways that culture, technology, and geography interact to determine whether a country can be caught in the poverty trap of our model.
10. What is unconditional aid? Explain the effect of unconditional aid on worker productivity growth in the poor country.
11. Explain why opening the economy to international capital flows does not necessarily equate k across countries.
12. Explain why opening the economy to international capital flows alters the economy’s fiscal policy? Explain why the new fiscal policy is more pro-growth.
13. What is the worker productivity gap between rich and poor countries after the economy is opened? Explain why the gap is different than in the closed economy case.
14. Do all generations of households benefit from opening the economy to capital flows? Does the government benefit?
15. How can a program such as *Progresa* eliminate the poverty trap?
16. Compare the effects of the *Progresa* program to compulsory schooling.
17. Describe the fiscal policy reform experiment? What are the effects on worker productivity?
18. What is the aid cost of reform? For which policies is the aid cost the lowest? The highest?
19. Why does aid often fail to generate economic growth?
20. Can humanitarian aid hurt economic growth? Explain.
21. Explain *G4*.

Problems

1. Solve the household maximization problem to derive (5.1).
2. Derive the transition Eq. (5.6) for private physical capital intensity. Explain why private physical capital accumulation is positively affected by public capital and negatively affected by taxation.
3. Give a clear and intuitive interpretation of the government objective function (5.7).
4. Follow the steps in the Appendix to derive (5.10).
5. Use (5.4a) and Table 5.2 to compute the ratio of the marginal products of capital across the rich and poor countries, i.e. compute r^P/r^R. If the annualized value of the marginal product is 7% in the rich country, what is the predicted annualized marginal product in the poor country?

   Caselli and Feyrer (2007) compute relative marginal products across countries. Their "naïve" estimate of the ratio of the marginal products across rich and poor countries is quite similar to the ratio you just computed. However, they claim that the ratio is too large because the relative price of physical capital goods is much higher in poor countries than in rich countries, lowering the relative return to purchasing and using capital there (one sector models do not capture this because the relative price of capital and consumer goods is identically one). In our model what must change about the ratios of g and k across the rich and poor countries to reduce r^P/r^R?
6. Derive an expression for private capital intensity when the poor country is open to international capital flows. Show that the equalization of the return to capital across countries in an open economy does not generally imply an equalization of private capital intensity.
7. Derive the schooling and fertility demand equations when the government subsidizes the forgone earnings of older children. Explain why the subsidy has an ambiguous effect on fertility.
8. Suppose the government gives parents a lump transfer v that is not tied to the forgone earnings of children. Solve the model for schooling and fertility in the presence of the lump sum transfer. How are schooling and fertility affected by the transfer? Are the effects different than under the subsidy in problem 7?

Appendix

Alternative Interpretations of the Cost of Children

In the household model of fertility the cost of children can be thought of as **lost adult consumption** needed to feed children or **forgone work time** needed to raise children. In reality, both costs are important but it is simpler in the model to assume just one type of cost. In this chapter we think of the cost of children as representing lost adult consumption that is proportional to parent's wages because, especially in developing countries, we think of this as the most important cost quantitatively. We use the cost of time rearing in the second half of the book because it gives rise to simpler expressions in the theory and we are not as concerned about quantitative accuracy as we are here.

The key difference between the two cost interpretation surfaces when you calculate the economy's supply of labor. If the cost of children is lost adult consumption then the family supply of effective labor is

$$\widehat{h}_t \equiv h_t + n_{t+1}\gamma\bar{h}(T - e_t). \qquad \text{(lost consumption interpretation)}$$

Alternatively, under the forgone work time interpretation we have

$$\widehat{h}_t \equiv h_t(1 - \eta n_{t+1}) + n_{t+1}\gamma\bar{h}(T - e_t). \qquad \text{(forgone work time)}$$

Note that the labor supply of the adult worker is reduced by the time spent raising children. Using (5.1a), we can simplify the expression for human capital per family as.

$$\widehat{h}_t = h_t \frac{1+\beta}{1+\beta+\psi}. \qquad \text{(forgone work time)}$$

Optimal Fiscal Policy in a Closed Economy

Domestic fiscal policy is determined by maximizing (5.7) subject to the government budget constraint and the accumulation equations for private and public capital. The private household's indirect utility function may be written as.

$$U_t = U_0 + \bar{U}_t + (1+\beta)\ln\left((1-\tau_t)w_t D_t\right) + \beta \ln R_t + \psi \ln\left((1-\tau_{t+1})w_{t+1}D_{t+1}\right),$$

where U_0 is a constant and $\bar{U}_t = (1+\beta)\ln h_t + \psi \ln n_{t+1} + \psi \ln h_{t+1}$ is independent of fiscal policy. For the purpose of setting optimal fiscal policy, the government can then be modeled as choosing tax rates and public capital to maximize,

$$\sum_{t=0}^{\infty} \beta^t \left(\ln c_t^g + \phi\{(1+\beta)\ln\left((1-\tau_t)w_t D_t\right) + \beta \ln R_t + \psi \ln\left((1-\tau_{t+1})w_{t+1}D_{t+1}\right)\}\right) \tag{5.7'}$$

subject to (5.4), (5.6), (5.8), and (5.9).

Substituting the constraints into the objective function and collecting common terms yields the following equivalent problem

$$\max_{\{\tau_{t+1},g_{t+1},k_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty}\beta^t \ln\left[\tau_t k_t^{\alpha} g_t^{\mu(1-\alpha)}\widehat{h}_t - g_{t+1}(1+q)n_{t+1}\right]$$
$$+\phi\sum_{t=0}^{\infty}\beta^t\{[\beta(\alpha-1)+\psi\alpha+\beta\alpha(1+\beta)]\ln k_{t+1}$$
$$+\mu(1-\alpha)[(\beta+\psi)+\beta(1+\beta)]\ln g_{t+1}+[\beta+\psi+\beta(1+\beta)]\ln(1-\tau_{t+1})\}$$
$$+\sum_{t=1}^{\infty}\lambda_t\left\{\left[\frac{\beta}{1+\beta+\psi}\right]\frac{(1-\tau_{t-1})(1-\alpha)k_{t-1}^{\alpha}g_{t-1}^{\mu(1-\alpha)}h_{t-1}}{(1+q)n_t\widehat{h}_t}-k_t\right\},$$

where λ is the multiplier associated with the private capital accumulation constraint.

To solve this sequence problem, begin by differentiating to get the first-order conditions for $\tau_t, g_t, k_t, \lambda_t$, for $t \geq 1$. Be careful to differentiate wherever the choice variable appears in the objective function. Next, substitute into the first order conditions the "guess" $(1+q)n_{t+1}g_{t+1} = B\tau_t k_t^{\alpha} g_t^{\mu(1-\alpha)}\widehat{h}_t$, where B is an undetermined coefficient. Finally, solve the first order conditions for B, τ_t, g_t, and k_t to get (5.10).

A tricky part of the solution given by (5.10) involves the first order condition for k_t. This equation, along with the first order condition for λ_t and the guess for the g_t, can be used to solve for the expression $\lambda_t k_t$ by solving the following difference equation, $\lambda_t k_t = \beta^{t-1}\left\{\frac{\alpha\beta}{1-B}+\phi[\beta(\alpha-1)+\alpha\psi+\alpha\beta(1+\beta)]\right\}+\alpha\lambda_{t+1}k_{t+1}$, to get $\lambda_t k_t = \frac{\beta^{t-1}}{1-\alpha\beta}\left\{\frac{\alpha\beta}{1-B}+\phi[\beta(\alpha-1)+\alpha\psi+\alpha\beta(1+\beta)]\right\}$. Use this solution to eliminate $\lambda_{t+1}k_{t+1}$ in the first order conditions for τ_t and g_t. Using the guess for g_t, these two first order conditions can be used to solve for τ_t and B to get $\tau_t = \tau = \frac{1-\alpha\beta}{1+(1-\alpha\beta)(1-B)\phi\Gamma}$ and $B = \frac{\frac{\mu\beta(1-\alpha)}{1-\alpha\beta}+\mu\beta(1-\alpha)\phi\Gamma}{1+\mu\beta(1-\alpha)\phi\Gamma}$. Combining these two expressions, completes the solution.

Optimal Fiscal Policy in an Open Economy

In an open economy, the government's problem can be written so that it solves

$$\max_{\{\tau_{t+1},g_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty}\beta^t \ln\left[\tau_t\left(\frac{(1-\tau_t)\alpha}{r*}\right)^{\frac{\alpha}{1-\alpha}} g_t^{\mu}\widehat{h}_t - g_{t+1}(1+q)n_{t+1}\right]$$
$$+\phi[\psi+\beta(1+\beta)]\sum_{t=0}^{\infty}\beta^t\left\{\frac{1}{1-\alpha}\ln(1-\tau_{t+1})+\mu\ln g_{t+1}\right\}.$$

This problem differs from the closed economy problem because private capital intensity is now determined by international capital flows rather than domestic saving. In a closed economy, government policy affected private capital formation by affecting the after-tax wage of savers that fund the subsequent period's private capital intensity. Now government policy affects private capital intensity by affecting the marginal product of private investments in the poor country—reduced by higher tax rates and raised by higher public capital intensity. In an open economy, government policy has a more *immediate* effect on private capital formation—this period's policy affects *this* period's capital intensity rather than this period's saving flow and *next* period's capital intensity.

Differentiating with respect to τ_{t+1} and g_{t+1} generates first order conditions. As before guess a solution for g of the form.

$$(1+q)n_{t+1}g_{t+1} = B\tau_t\left(\frac{\alpha(1-\tau_t)}{r*}\right)^{\frac{\alpha}{1-\alpha}} g_t^{\mu}\widehat{h}_t.$$

Substitute into the first order conditions and solve for τ_{t+1} and B to get the solution in the text.

References

Acemoglu D (2009) Introduction to modern economic growth. Princeton University Press, Princeton, New Jersey.

Acemoglu D, Robinson J (2012) Why nations fail: The origins of power, prosperity, and poverty. Crown Publishers, New York.

Aschauer D (1989) Is public expenditure productive? Journal of monetary economics 23 (2):177–200.

Azariadis C, Drazen A (1990) Threshold externalities in economic development. Quarterly journal of economics 105(2):501–526.

Azariadis C, Stachurski J (2004) Poverty traps. In: Aghion P, Durlauf S (eds) Handbook of economic growth. Amsterdam, North Holland.

Becker G, Murphy K, Tamura R (1990) Human capital, fertility, and economic growth. Journal of political economy 98(5):S12–S37.

Bongaarts J (2002) The end of the fertility transition in the developing world. Population and development review 28(3):419–443.

Caselli, F. and Fyrer, J., 2007, The Marginal Product of Capital, *Quarterly Journal of Economics*, 122, 535–568.

Clarida RH (1993) International capital mobility, public investment, and economic growth. NBER working paper 4506. National Bureau of Economic Research, Cambridge, Massachusetts.

Córdoba J, Ripoll M (2013) What explains schooling differences across countries? Journal of monetary economics 60(2):184–202.

de la Croix D (2013) Fertility, education, growth, and sustainability. Cambridge University Press, Cambridge, UK.

Doepke M (2004) Accounting for fertility decline during the transition to growth. Journal of economic growth 9(3):347–383.

Drazen A (1978) Government debt, human capital, and bequests in a life-cycle model. Journal of political economy 86(6):505–516.

Drazen A (2000) Political economy in macroeconomics. Princeton University Press, Princeton, New Jersey.

Drazen A, Isard P (2004) Can public discussion enhance program ownership? NBER working paper 10927. National Bureau of Economic Research, Cambridge, Massachusetts.

Easterly W (2001) The elusive quest for growth. MIT Press, Cambridge, Massachusetts.

Easterly W (2007) The white man's burden. Penguin Press, New York.

Easterly W, Levine R, Roodman D (2004) Aid, policies, and growth. American economic review 94 (3):774–780.

Galor O (2011) Unified growth theory. Princeton University Press, Princeton, New Jersey.

Galor O, Moav O (2002) Natural selection and the origins of economic growth. Quarterly journal of economics 117(4):1133–1192.

Galor O, Ziera J (1993) Income distribution and macroeconomics. Review of economic studies 60 (1):35–62.
Glomm G, Ravikumar B (1997) Productive government expenditures and long-run growth. Journal of economic dynamics and control 21(1):183–204.
Grossman G, Helpman E (2001) Special interest politics. MIT Press, Cambridge, Massachusetts.
Hall R, Jones C (1999) Why do some countries produce so much more output per worker than others? Quarterly journal of economics 114(1):83–116.
Isard P, Lipschitz L, Mourmouras A, Yontcheva B (2006) High-level seminar on foreign aid and macroeconomic management: Overview and synopsis of key lessons. In: Isard P, Lipschitz L, Mourmouras A, Yontcheva B (eds) The macroeconomic management of foreign aid. International Monetary Fund, Washington, DC.
Ivanyna, M, Mourmouras, A and Rangazas, P (2018) The macroeconomics of corruption: governance and growth, Springer, New York.
Krueger A (2002) A model for evaluating the use of development dollars, south of the border. New York Times, May 2.
Lord W (2002) Household dynamics, policy, and economics growth. Oxford University Press, Oxford, UK.
Lord W, Rangazas P (2006) Fertility and development: The roles of schooling and family production. Journal of economic growth 11(3):229–261.
Mallaby S (2004) The world's banker. Penguin Press, New York.
Marchesi S, Sabani L (2007) IMF concerns for reputation and conditional lending failure: Theory and empirics. Journal of development economics 84(2):640–666.
Mourmouras A, Mayer W (2005) The political economy of IMF conditionality: A common agency model. Review of development economics 9(4):449–466.
Mourmouras A, Rangazas P (2007) Foreign aid policies and sources of poverty: a quantitative framework. IMF staff papers 54: 59–90.
Nunn N, Qian N (2014) US food aid and civil conflict. American economic review 104 (6):1630–1666.
Persson T, Tabellini G (1990) Macroeconomic policy, credibility, and politics. Harwood, London.
Poterba J (1997) Rate of return to corporate capital and factor shares. NBER working paper 6263. National Bureau of Economic Research, Cambridge, Massachusetts.
Pritchett L (2000) The tyranny of concepts: CUDIE (cumulated depreciated investment effort) is not capital. Journal of economic growth 5(4):361–384.
Radelet S, Clemens M Bhavnani R (2006) Aid and growth: The current debate and some new evidence. CGD working paper 133. Center for Global Development, Washington, DC.
Raghuram R, Subramanian A (2008). Aid and growth: What does the cross-country evidence really show? Review of economics and statistics 90(4):643–665.
Rajan R, Subramanian A (2011) Aid, Dutch disease, and manufacturing. Journal of development economics 94(1):106–118.
Rangazas P (2000) Schooling and economic growth: A King-Rebelo experiment with human capital. Journal of monetary economics 46(2):397–416.
Scholl A (2009) Aid effectiveness and limited enforcement conditionality. Review of economic dynamics 12(2):377–391.
Svensson J (2000) When is foreign aid policy credible? Aid dependence and conditionality. Journal of development economics 61(1):61–84.
Weil D (2009) Economic growth. Addison Wesley, New York.

Part II

Two-Sector and Dual Economies

Two Sector Growth Models

6

This section provides an introduction to two sector growth models. We begin with a model where a single good is produced using traditional means of production. In the *traditional* sector, production is carried out by households using land (natural resources) and labor. There are no firms or factories that rely on heavy plant and equipment and modern production methods to produce goods. This setting can be used to identify the conditions necessary for a *modern* sector to appear that would begin an "industrial revolution," as in Hansen and Prescott (2002), helping to explain *G8* and *G9*.

Next, we examine a two sector model where two distinct types of goods are produced—agricultural and industrial. The model is first analyzed under the assumption that the economy is perfectly closed to the international trade of goods. An open economy version is then considered. An important task is to determine the circumstances under which opening the economy to the international trade of goods increases growth and welfare in the long run. The answer can be quite different from the typical static analysis of international trade, wxhere trade is shown to be welfare improving for all countries involved.

Finally, we extend the model to include health benefits from food consumption. The extended model is used to explain several facts about developing economies including the approximate constancy of caloric intake, the rise in body mass and health, and the declining budget share devoted to food referred to in *G10*. This material is related to the "subsistence constraint," a level of consumption required for survival, a frequent feature of development models. One example is the Malthusian model from Chap. 4. In the literature, the subsistence constraint plays a role in explaining a wide variety of development facts including rising saving rates, initially rising fertility rates, and falling employment shares in agriculture.

S. Das et al., *Economic Growth and Development*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-319-89755-4_6

6.1 From Stagnation to Growth

For most of human history there was virtually no sustained increase in the standard of living of the average person (Clark (2007, Part I) and Galor (2011, Chapter 2)). Living standards are closely connected with output or income per person—also known as per capita income. Before 1700, per capita income was stagnant across the world (Galor (2005, Fig. 2.1)). England began to see some sustained increases in per capita income during the eighteenth century, but the growth rates in per capita income were modest, certainly less than 1% per year (Crafts (1995, Table 1)). Before 1800, the growth rate in per capita income in Western Europe as a whole was barely above one half percent (Galor and Weil (2000, Figure 1)). In the U.S., growth rates in per capita income were close to zero before 1800 (Lucas (2002, Table 5.2)). The lack of significant sustained growth in any particular region meant that living standards did not differ dramatically *across* regions. In 1820, Western Europe had per capita that was 1.7 times higher than in Latin America, 2.1 times higher than Asia, and 2.9 times higher than Africa (Galor (2005, section 2.1.1)).

After 1800, the nature of economic growth changed. The modest growth in England accelerated and spread throughout Western Europe. Income per capita grew between 1.5% and 2.0% in Western Europe from 1820 to 1929 (Galor and Weil (2000, Figure 1)). In the U.S., growth in income per capita and output per worker exceeded 1.5% by the middle of the nineteenth century and then remained between 1.5% and 2.5% throughout the twentieth century (Lucas (2002, Table 5.2) and Mourmouras and Rangazas (2009 Table 2)).

Not all countries began modern growth in the nineteenth century. As a result, the income gaps between countries began to grow—a phenomenon known as the *Great Divergence*. Over a period of two centuries, Parente and Prescott (2000) compare per capita incomes across large Eastern countries and Western countries.[1] As mentioned, income gaps were not large at the beginning of the nineteenth century. At this time, per capita income in the West was only 2.1 times higher than in the East. By 1950 the income gap widened considerably, with per capita in the West becoming 7.5 times higher than in the East. More dramatic gaps are found when comparing the richest and poorest countries of the world. A narrower set of Western offshoots, defined by Maddison (1995), comprised of the United States, Canada, Australia, and New Zealand, formed the richest set of countries in the world by the middle of the twentieth century. The per capita income of these countries in 1950 was 15 times higher than those in Africa (Galor (2005)). By 2000, rich country per capita income was 18 times higher.

Modeling the sources and timing of the industrialization and the onset of modern growth is important in explaining income differences across countries today. The primary focus of this section is to begin thinking about this issue. To do so, we

[1]The East consists of China, Pakistan, India, Bangladesh, Indonesia, Japan, Burma, the Philippines, South Korea, Taiwan, and Thailand. The West consists of Western European countries and their ethnic offshoots: Canada, Mexico, USA, Argentina, Brazil, Chile, Australia, and New Zealand.

initially assume that production is highly dependent on land and natural resources and think about the conditions needed for the traditional economy to transform into a modern economy where production relies heavily on man-made physical capital.

6.1.1 A Traditional Economy

During the pre-industrial period, production was carried out by households on "family farms" and other small-scale businesses. To present a sharp contrast to the modern capital-intensive production that provides the basis of standard neoclassical theory, we assume that pre-industrial traditional production uses only land (L) and "farm" labor (F) as inputs. The traditional sector production function is given by

$$O_t = \tilde{A} L_t^{\alpha} \tilde{H}_t^{1-\alpha}, \tag{6.1}$$

where O denotes output produced in the traditional sector, $\tilde{A}$ is TFP in the traditional sector, and $\tilde{H} = \tilde{D}F$ is the effective workforce, indexed by the state of technology, $\tilde{D}$, in the traditional sector.[2]

Throughout this chapter we assume that there are perfectly competitive markets for both labor and land. The extent to which these markets are well-functioning influences the value of $\tilde{A}$. Effective labor can be hired at the competitive wage, $\tilde{w}$, and land can be rented at the competitive rental rate for land, r^L. The profit-maximizing use of effective labor and land by traditional sector producers requires marginal products be equated to factor prices,

$$\tilde{w}_t = (1-\alpha) \tilde{A} L_t^{\alpha} \tilde{H}_t^{-\alpha} \tag{6.2a}$$

$$r_t^L = \alpha \tilde{A} L_t^{\alpha-1} \tilde{H}_t^{1-\alpha}. \tag{6.2b}$$

There are N_t young households in each period. Households live for two periods. In the first period, each household supplies one unit of labor and earns the wage, $\tilde{w}\tilde{D}$. Part of the wage is consumed, c_1, and part is saved by purchasing land, l, at the competitive land price, p^L. In the second period households retire and consume goods, c_2, financed by the rent they earn on land and the proceeds from selling land.

A generation-t household chooses consumption and saving to maximize the utility function

$$U_t = \ln c_{1t} + \beta \ln c_{2t+1}$$

subject to the two single period budget constraints

[2]Human capital and fertility are treated exogenously in this chapter. We add endogenous schooling and fertility back into the model in Chap. 7.

$$c_{1t} + p_t^L l_t = \tilde{w}_t \tilde{D}_t$$

and

$$c_{2t+1} = \left(r_{t+1}^L + p_{t+1}^L\right) l_t.$$

Combining the two single period budget constraints gives the lifetime budget constraint.

$$c_{1t} + \frac{c_{2t+1}}{R_t^L} = \tilde{w}_t \tilde{D}_t,$$

where the return to purchasing one unit of land in period t is $R_t^L \equiv \dfrac{p_{t+1}^L + r_{t+1}^L}{p_t^L}$. The demand functions that result from solving the maximization problem are

$$c_{1t} = \frac{\tilde{w}_t \tilde{D}_t}{1+\beta} \tag{6.3a}$$

$$c_{2t+1} = R_t^L \frac{\beta \tilde{w}_t \tilde{D}_t}{1+\beta} \tag{6.3b}$$

$$p_t^L l_t = \frac{\beta \tilde{w}_t \tilde{D}_t}{1+\beta}. \tag{6.3c}$$

These demand functions are similar to those encountered in Chap. 2. Consumption and saving, now in the form of land purchases, are fractions of wages that depend on the household's time preference.

Market clearing in the labor market requires that the demand for labor by traditional producers equals the supply of labor from households, $F_t = N_t$. Market clearing in the land market requires that the households' demand for land as an asset equals the fixed stock of available land, $L_t \equiv N_t l_t = L$. Using the market clearing conditions, and assuming exogenous growth factors for technology $(1 + \tilde{d})$ and the population (n), the equilibrium price and return to land are

$$p_t^L = \frac{\beta \tilde{w}_t \tilde{D}_t}{1+\beta} \frac{N_t}{L} \tag{6.4a}$$

$$R_t^L = \left[1 + \frac{\alpha(1+\beta)}{\beta(1-\alpha)}\right] \left[\left(1+\tilde{d}\right) n\right]^{1-\alpha}. \tag{6.4b}$$

Equation (6.4a) indicates that the price of land rises as the economy grows and the demand for the fixed supply of land increases. Equation (6.4b) tells us that the return to land is constant. This is because the growth rate of land prices is constant and the growth in the marginal product of land, and the rental rate for land, is the same as the growth rate in land prices. The higher is the growth in effective labor supply $((1+\tilde{d})n)$, the higher is the return to land.

As indicated previously, the pre-modern growth period was characterized by very slow, if any, sustained growth in labor productivity and population. For now, we treat both these characteristics as exogenous and for simplicity set $\tilde{d}=0, \tilde{D}=1$, and $n=1$. In the pre-modern steady state we then have

$$\tilde{w}=(1-\alpha)\tilde{A}\tilde{l}^{\alpha} \tag{6.5a}$$

$$R^{L}=1+\frac{(1+\beta)\alpha}{\beta(1-\alpha)}, \tag{6.5b}$$

where $\tilde{l}$ is land relative to the *effective* work force.

6.1.2 Onset of Modern Growth

Now think of the potential appearance of firms and a modern sector of production. The key task is to establish the conditions needed for the modern production to be profitable. Firms will have to pay workers at least what they can make in the traditional sector and capital owners at least what they can make by owning land.

As in the one sector models we have studied, the technology of the modern sector is given by

$$Y_t=AK_t^{\alpha}H_t^{1-\alpha}, \tag{6.6}$$

where $H_t=D_tM_t$. Firms must be able to generate nonnegative economic profit,

$$Y_t-w_tD_tM_t-r_tK_t\geq 0 \tag{6.7}$$

when paying workers what they could earn in the traditional sector, $w_tD_t=\tilde{w}\tilde{D}$, and capital owners what they could earn from owning land, $R_{t-1}\equiv 1+r_t-\delta=R^L$ or $r_t=\delta+\frac{(1+\beta)\alpha}{\beta(1-\alpha)}$.

To uncover the conditions under which (6.7) is satisfied, we need to relate (6.7) to the fundamental features of the economy. Begin by pulling M_t out of the expression in (6.7) and write the right-hand-side in terms of K_t/M_t. Using the first order condition for the profit-maximizing choice of K_t, $\alpha AD_t^{1-\alpha}(K_t/M_t)^{\alpha-1}=r_t$, write K_t/M_t in terms of r_t. Finally, substitute $w_tD_t=\tilde{w}\tilde{D}$ and $r_t=\delta+\frac{(1+\beta)\alpha}{\beta(1-\alpha)}$ into (6.7) and rearrange to get

$$D_t\geq\tilde{A}\tilde{l}^{\alpha}\left(\frac{\delta}{\alpha}+\frac{1+\beta}{\beta(1-\alpha)}\right)^{\frac{\alpha}{1-\alpha}}A^{\frac{-1}{1-\alpha}}. \tag{6.8}$$

The condition in (6.8) gives the threshold value that technological progress in the modern sector must exceed before production becomes profitable at the factor prices paid in the traditional sector. Until technology (D_t) has developed sufficiently, the modern sector will not appear. This raises the question of what caused D_t to grow and eventually reach the threshold in (6.8).

We take technology as exogenous and thus cannot explain why technological progress eventually caused the threshold in (6.8) to be exceeded, opening the way for sustained modern growth. Acemoglu and Robinson (2012), Clark (2007), and Galor (2011) provide theories of the early progress in D_t. All the theories argue that technological advances occurred very slowly before 1800. Important determinants of the early progress are believed to be the slow growth in population size (more people, more ideas), institutional change that encouraged property rights and innovation, and genetic transmission of the traits that make humans productive (richer people had bigger families than poorer people).

Even if the state of technology is common knowledge to the world as a whole, the threshold may nevertheless be met in some countries but not in others. Differences in thresholds across countries led to differences in the onset of modern growth, causing standard of livings to diverge beginning around 1800. The technological threshold is higher for countries with plentiful and productive land or natural resources ($\tilde{A}\tilde{l}^{\alpha}$), a type of "natural resource curse." For example, Habbakuk (1962) argued that United States labor productivity in traditional agriculture was high in the early nineteenth century, due to the abundance of land, and as a result slowed the structural transformation of the economy. The availability of land and natural resources in Africa has often been cited as one reason that its take-off to a sustained modern growth path has been delayed.

Parente and Prescott (2000) stress the importance of political forces that restrict work practices, use of machinery, and adoption of the available technology, all making A low and increasing the threshold for take-off. The most famous historical example of such forces were the Luddites, named after one of their leaders (Ned Ludd), nineteenth-century English textile artisans who violently protested against the machinery introduced during the early stages of the Industrial Revolution.

A lack of institutions that support markets, such as property right protection or public infrastructure, lowers TFP in each sector. However, from (6.8), we see that any economy-wide deficiencies that lower A and $\tilde{A}$ proportionally will *disproportionately* harm the modern sector. This is because a lower value of A lowers the demand for physical capital and profits, for given factor prices. A low value of A causes the marginal product of capital to be low, resulting in low capital demand, low production, and low profits in the modern sector. There is no similar indirect effect of TFP in the traditional sector because land is fixed and therefore is not responsive to the level of $\tilde{A}$. Thus, poor support for markets, generally throughout the entire economy, can prevent the onset of modern growth by disproportionately harming the modern sector through its negative effect on capital accumulation.

6.2 The Structural Transformation of a Two-Sector Economy

Our focus now turns to the rate of growth of an economy after the threshold condition given by (6.8) is met and a modern sector appears. In other words, we examine what determines the pace of the structural transformation from a traditional economy to a modern economy once both sectors are operating.

As suggested in the previous section, if two different sectors are to operate, they must pay workers the same wage and asset owners the same return,

$$w_t D_t = \tilde{w}_t \tilde{D}_t \tag{6.9a}$$

$$R_t = R_t^L, \tag{6.9b}$$

where we now allow for technological progress in the traditional sector.

Households can work in either sector. Households have the portfolio choice of saving by purchasing physical capital or land, so their budget constraints become

$$c_{1t} + p_t^L l_t + s_t^K = w_t D_t$$

and

$$c_{2t+1} = \left(r_{t+1}^L + p_{t+1}^L\right) l_t + R_t s_t^K,$$

where $p_t^L l_t + s_t^K$ represents total saving and s_t^K is saving through purchases of physical capital. Note that the budget constraints apply to all households, no matter which sector they are employed. This is because the wage is the same in either sector, and any household can purchase land or capital in national markets.

We seek a transition equation for physical capital intensity, so we need to start with an expression for s_t^K. Using (6.9b), the household's optimization problem can be solved to get the following expression for the household's supply of capital

$$s_t^K = \frac{\beta}{1+\beta} w_t D_t - p_t^L l_t \tag{6.10}$$

As in Chap. 2, equilibrium in the capital market requires that the capital demanded by firms for production in period $t+1$ be supplied by household purchases of capital in period t, $K_{t+1} = N_t s_t^K$. We define k to be the capital to effective labor ratio in the modern sector, $k_{t+1} = K_{t+1}/H_{t+1}$, where $H_{t+1} = M_{t+1} D_{t+1}$. The number of workers in the modern sector is now a fraction π_{t+1} of the entire supply of labor, $M_{t+1} = \pi_{t+1} N_{t+1}$.

Using the same approach as in previous chapters, the capital market equilibrium condition is used to derive the transition equation for k as

$$k_{t+1} = \frac{(1-\alpha) A k_t^\alpha}{\pi_{t+1} n (1+d)} \left[\frac{\beta}{1+\beta} - \frac{p_t^L l_t}{(1-\alpha) A k_t^\alpha D_t}\right]. \tag{6.11}$$

Comparing (6.11) to the transition equation for the basic one sector model, we see there are two differences. First, the number of workers in the modern sector is now endogenous, even if population growth is exogenous. Migration of workers from the traditional to the modern sector crowds the available capital and lowers k. Second, land holdings offer households an alternative way to save. The larger are the value of land holdings, relative to the household wages, the less saving is devoted to physical capital and the smaller is k. To use (6.11) to solve for capital accumulation, we must close the model by determining the paths of π_t and p_t^L.

6.2.1 Labor Market Equilibrium

The equilibrium allocation of labor across sectors is determined by the condition that the wage is the same in each sector. The flow of labor across sectors forces wages across sectors to equalize. Writing out (6.9a) in terms of the marginal products of labor and using the fact that $\tilde{H}_t = F_t\tilde{D}_t = (1-\pi_t)N_t\tilde{D}_t$, yields the following expression for the fraction of the work force in the traditional sector that equilibrates wages

$$1 - \pi_t = \left(\frac{\tilde{A}\tilde{D}_t}{AD_t}\right)^{\frac{1}{\alpha}}\frac{\tilde{l}_t}{k_t}, \tag{6.12}$$

where $\tilde{l} \equiv L/N\tilde{D}$ is now interpreted as the ratio of land to the maximum potential effective labor supply in the traditional sector.

The relative size of the traditional sector's work force is determined by the relative productivity of labor in the traditional sector. The relative productivity of traditional labor is determined by the relative state of efficiency and technology in the traditional sector and by its relative factor endowment. The traditional sector shrinks faster, the faster is the growth in relative technological progress in the modern sector and the faster is physical capital accumulation. Even if (i) technological progress is uniform across sectors and (ii) capital intensity in the modern sector remains constant, the traditional sector will shrink because land becomes scarce relative to the economy's effective labor supply, provided there is positive population growth or positive technological progress.

6.2.2 Land Market Equilibrium

To determine the equilibrium price of land begin with the condition that the return to owning land must equal the return to owning physical capital,

$$\frac{p_{t+1}^L + r_{t+1}^L}{p_t^L} = 1 + r_{t+1} - \delta.$$

Using the profit maximizing conditions for renting land and capital, along with (6.12), one can derive the following difference equation in p^L,

$$p_{t+1}^L = \left(1 + \alpha A k_{t+1}^{\alpha-1} - \delta\right) p_t^L - \alpha \tilde{A} k_{t+1}^{\alpha-1} \left(\tilde{a} \frac{\tilde{D}_{t+1}}{D_{t+1}}\right)^{\frac{1-\alpha}{\alpha}} \tag{6.13}$$

where $\tilde{a} \equiv \tilde{A}/A$.

The dynamics of the economy is solved by first substituting (6.12), dated for $t+1$, into (6.11) to eliminate π_{t+1}. Equations (6.11) and (6.13) then form a system of first order difference equations in k and p^L. We now have a two dimensional system rather than the single difference equation that generates the dynamics in a one sector model. Moreover, there is an important additional difference. We do not have an initial condition that pins down the initial price of land. In this situation it is possible to have multiple equilibrium paths for the economy, each corresponding to a different initial value for the price of land. It is also possible that there is a unique price of land that is consistent with convergence to the economy's steady state. In this case, a *terminal condition*, convergence to the steady state, replaces an initial condition in determining the initial price of land. Whether or not there is a unique stable equilibrium path depends on the parameter values chosen for the model. We will address the conditions under which there is a unique equilibrium path in the next section. In this section we side-step a complete dynamic analysis by focusing only on the steady state of the economy.

6.2.3 Steady State Equilibrium

With both a traditional and a modern sector producing the *same* goods, it is natural to examine the conditions where the traditional sector gradually disappears and the economy converges to a steady state where the goods are produced in the modern sector only. To begin with, the natural assumptions are that (i) technological progress is faster in the modern sector than in the traditional sector ($d > \tilde{d}$) and (ii) population growth is positive ($n > 1$). Under these assumptions, Eq. (6.12) tells us that if the economy converges to a given value of k, then the fraction of the work force employed in the traditional sector must go to zero. Furthermore, the rental rate on land can be written as

$$r_{t+1}^L = \alpha \tilde{A} k_{t+1}^{\alpha-1} \left(\tilde{a} \frac{\tilde{D}_{t+1}}{D_{t+1}}\right)^{\frac{1-\alpha}{\alpha}}.$$

Thus, if the economy converges to a given value of k and $d > \tilde{d}$ then land rents must also go to zero.

Now, with land rents going to zero, one would expect that the price of land would also go to zero. Unfortunately, things are not this simple. It is possible for an asset that has no fundamental value, i.e. an asset that is not productive, to have a positive price. For land to have a positive equilibrium price in the steady state without a traditional sector, it is still the case that the two dynamic Eqs. (6.11) and (6.13) must be satisfied.

For there to be a steady state k that satisfies (6.11) with $\pi = 1$, the term $p_t^L l_t / D_t$ must remain constant. This implies land prices must appreciate at a rate satisfying

$$\frac{p_{t+1}^L}{p_t^L} = (1 + d)n. \tag{6.14}$$

Label the steady state value of k that solves (6.11), when (6.14) is also satisfied, $\bar{k}$. Next, for (6.13) to be satisfied, with land rents going to zero, land prices must appreciate so that

$$\frac{p_{t+1}^L}{p_t^L} = 1 + \bar{r} - \delta, \tag{6.15}$$

where $\bar{r}$ is the interest rate associated with $\bar{k}$. Combining (6.14) and (6.15), gives

$$1 + \bar{r} - \delta = (1 + d)n. \tag{6.16}$$

There could be a configuration of parameters where the $\bar{k}$ satisfying (6.16) also satisfies (6.11) because there are quite a few parameters in (6.11) that are not in (6.16). If both (6.11) and (6.16) are satisfied at $\bar{k}$, we would have a steady state with a permanent asset "bubble," i.e. land is continually appreciating despite the fact that land has no fundamental value.

To think about the conditions needed to eliminate the possibility of a steady state bubble, note that if land prices go to zero, the steady state would be characterized by

$$k = \left[\frac{(1 - \alpha)A\beta}{n(1 + d)(1 + \beta)}\right]^{\frac{1}{1-\alpha}} \tag{6.17a}$$

$$r = \left[\frac{\alpha(1 + \beta)}{(1 - \alpha)\beta}\right] n(1 + d). \tag{6.17b}$$

If α is about 1/3, as the evidence indicates, and $\beta < 1$ (future utility is discounted relative to current utility), then (6.17b) indicates $r > n(1 + d)$. Next notice from (6.11) that $\bar{k} < k$ and therefore $\bar{r} > r > n(1 + d)$. This means that (6.16) is not satisfied because the return to capital, $1 + \bar{r} - \delta$, is strictly greater than the return to land, $n(1 + d)$, if $\delta \leq 1$. Thus, in this model, for reasonable parameter settings, bubbles are not possible. Eventually, capital dominates land as an asset and land prices go to zero in the long run. However, it is interesting to note that perfectly competitive models without uncertainty, or special assumptions about speculative behavior, can generate permanent asset bubbles.

6.2.4 Extensions to Human Capital and Fertility

As in the one sector case, the model can be easily extended to include schooling investments in human capital of children and an endogenous choice of fertility. We do this in Chap. 7 but would like to make one point here. Schooling offers an additional reason why an economy may move from stagnation to growth. Recall the possibility of a poverty trap from Chap. 4. More specifically, the theory of schooling from Chap. 4 is given by

$$e_t = \max\left[\frac{\theta\left(\eta\left(e_{t-1}/\bar{e}\right)^{\theta} - \gamma T\right)}{\gamma(1-\theta)}, \bar{e}\right] \tag{6.18}$$

Initially, the relative productivity of children, γ, may be sufficiently high that the optimal choice of schooling is $\bar{e}$. However, the appearance of child labor laws or compulsory schooling laws may reduce the employment opportunities for children and lower their relative productivity. If γ falls sufficiently, there will be a take-off of human capital accumulation over time. Thus, policies that lower the relative productivity of children can create a take-off, or at least an acceleration, of growth that compliments the advancement of technology in the modern sector.

Note also that one might give the traditional and modern sectors a geographic interpretation—the traditional sector is rural and the modern sector is urban. Suppose that to work in a given sector, the household must also live in that sector. Now think of an initial division of the population across the two sectors. Households initially born in the traditional sector might start with less schooling than those initially born in the modern sector, $\tilde{e}_t < e_t$ (perhaps because of a limited ability to enforce child labor laws in the traditional sector at some point in history or because of an absence of rural public schools). As indicated by the analysis in Chap. 4, the difference in schooling across sectors would also give rise to a difference in fertility, $n_t < \tilde{n}_t$.

If the underlying theory from Chap. 4 is used, then the differences in schooling and fertility are household-specific and will persist across generations. Thus, regardless of where the households choose to work and live, their past family history will dictate their fertility and schooling investments in their children. In future chapters, we will consider models where the family's locational choice can affect their fertility and human capital investments in children in this and other ways.

6.3 Two Sectors and Two Goods

Suppose now that there are two distinct goods. In the traditional sector, agricultural goods or primary products (O) are produced and in the modern sector, manufactured goods (Y) are produced. To simplify the notation and analysis a bit assume that the agricultural good is only consumed by young households, and that $n = 1, d = \tilde{d} = 0$, and $D_t = \tilde{D}_t = 1$. These assumptions allow us to focus on the transitional dynamics associated with physical capital accumulation and expansion of the modern sector in a relatively simple setting.

Household preferences are given by $U_t = u_{1t} + \beta u_{2t+1}$, with $u_{1t} = \chi \ln y_{1t} + (1-\chi) \ln o_{1t}$ and $u_{2t+1} = \ln y_{2t+1}$, and where χ is a preference parameter that weighs the relative utility received from manufacturing goods.

The single period budget constraints for the household are $y_{1t} + p_t o_{1t} + p_t^L l_t + s_t^K = w_t$ and $y_{2t+1} = \left(r_{t+1}^L + p_{t+1}^L\right) l_t + R_t s_t^K$, where p is the relative price of the agricultural good. Household demand for goods and assets are then

$$y_{1t} = \left[\frac{\chi}{1+\beta}\right] w_t \tag{6.19a}$$

$$y_{2t+1} = \left[\frac{\beta}{1+\beta}\right] R_t w_t \tag{6.19b}$$

$$o_{1t} = \left[\frac{1-\chi}{1+\beta}\right] \frac{w_t}{p_t} \tag{6.19c}$$

$$s_t^K = \frac{\beta}{1+\beta} w_t - p_t^L l_t. \tag{6.19d}$$

6.3.1 Labor Market Equilibrium

As in Sect. 6.2, the equilibrium allocation of labor across sectors is determined by the condition that wage is the same in each sector. This gives an equilibrium condition that is similar to (6.12),

$$1 - \pi_t = \left(p_t \tilde{a}\right)^{1/\alpha} \frac{l_t}{k_t}. \tag{6.20}$$

The new feature is that the relative price of agricultural goods affects the allocation of labor. The higher is p_t, the more labor is allocated to the traditional sector, other things constant.

6.3.2 Goods Market Equilibrium

The relative price of the agricultural good equilibrates the supply and demand for the agricultural product. The output of the traditional sector is $\tilde{A} L^\alpha ((1-\pi_t)N)^{1-\alpha}$. Using (6.20), the traditional sector output can be written as $\tilde{A} L \left(p_t \tilde{a}\right)^{(1-\alpha)/\alpha} k_t^{\alpha-1}$. From (6.19), the market demand for traditional sector output is $N \frac{1-\chi}{1+\beta} \frac{(1-\alpha) A k_t^\alpha}{p_t}$. The equilibrium relative price that equates these two expressions is

$$p_t = \left[\frac{(1-\chi)(1-\alpha)}{1+\beta}\frac{k_t}{\tilde{a}^{1/\alpha}l}\right]^{\alpha}. \tag{6.21}$$

The relative price of agricultural goods rises with capital accumulation because of rising demand for, and falling supply of, the agricultural good. Capital accumulation causes wages and the demand for agricultural goods to rise. The rise in wages in the modern sector bids workers away from the traditional sector and causes the supply of agricultural goods to fall. Both of these forces result in higher agricultural prices.

Note that substituting (6.21) into (6.20) yields

$$1-\pi_t = \frac{(1-\chi)(1-\alpha)}{1+\beta}. \tag{6.22}$$

Taking into account the effect of capital accumulation on the relative price of the traditional good, reveals that the fraction of the work force in the traditional sector is constant. As capital accumulates, it raises the marginal product of labor in the modern sector and wages across the economy, which reduces the demand for labor in the traditional sector. However, capital accumulation also increases the relative price of the agricultural good, which raises the demand for labor in the traditional sector. Under our assumptions, these two effects exactly offset and the fraction of labor employed in the traditional sector remains constant. This is an interesting result because it shows that, in a closed economy, capital accumulation alone does not necessarily generate a structural transformation away from agriculture and toward manufacturing.

6.3.3 Land Market Equilibrium

We now turn to the determination of the rental rate for land and the equilibrium price of land. As in Sect. 6.2, the rental rate for land must equal the value of the marginal product of land in production, $r_t^L = \alpha p_t \tilde{A}(l/(1-\pi_t))^{\alpha-1}$. Using (6.21) and (6.22), the rental rate on land can be written as

$$r_t^L = \frac{\alpha(1-\chi)(1-\alpha)Ak_t^{\alpha}}{(1+\beta)l}. \tag{6.23}$$

Capital accumulation drives up land rents by increasing wages and the price of agriculture goods. The higher the price of agricultural goods, the higher is the value of the marginal product of land and the higher are land rents.

Next, we move to the price of land as an asset. If both assets are held, the return to owning land must equal the return to owning physical capital,

$$\frac{p_{t+1}^L + r_{t+1}^L}{p_t^L} = 1 + r_{t+1} - \delta.$$

Using the profit maximizing conditions for renting land and capital, gives us the following difference equation in p^L that is similar to (6.13) in the one good case,

$$p_{t+1}^{L} = \left(1 + \alpha A k_{t+1}^{\alpha-1} - \delta\right) p_{t}^{L} - \frac{\alpha(1-\chi)(1-\alpha)A}{(1+\beta)l} k_{t+1}^{\alpha}. \tag{6.24}$$

Note that an increase in capital intensity lowers the value of the right-hand-side of (6.24), as the return to capital falls and land rents rise. This implies that future land prices must fall relative to current land prices to restore equality. Intuitively, the current price of land rises as demand for land increases due to falling returns to capital and rising land rents. This is different than (6.13) where capital accumulation has an ambiguous effect on land prices because, there, an increase in k *lowers* land rents (as workers are pulled out of the traditional sector when wages rise in the modern sector).

6.3.4 Transition Equation for Capital

The dynamic system is completed by deriving the transition equation for physical capital per worker. Using the same approach to deriving the transition equation as in Sect. 6.2 yields

$$k_{t+1} = \left(\frac{1}{1 - \frac{(1-\chi)(1-\alpha)}{1+\beta}}\right) \left[\frac{\beta A(1-\alpha)}{1+\beta} k_{t}^{\alpha} - p_{t}^{L} l\right]. \tag{6.25}$$

Similar to (6.11), the introduction of two distinct sectors and land has an ambiguous effect on capital intensity because there are two opposing changes to the one sector transition equation from Chap. 2. The first expression on the right-hand-side increases capital intensity. This expression appears because only a fraction of the workforce, a fixed fraction under our assumptions, works in the modern sector and uses capital to produce. Since k is defined as capital per worker in the modern sector, the fewer workers in the modern sector the higher the capital intensity, other things constant. The second change is captured by the second expression in the square brackets on the right-hand-side. This expression captures the fact that land diverts household saving away from capital purchases, causing capital intensity to decrease.

We will examine the transitional dynamics associated with (6.24) and (6.25), a topic for more advanced students, but before doing so we begin by discussing the steady state of the system.

6.3.5 Steady State Equilibrium

We can further simplify the dynamic analysis by assuming that $\delta = 0$. With no depreciation, the steady state version of (6.24) gives us the following expression for the steady state price of land

$$\bar{p}^L = \frac{(1-\chi)(1-\alpha)}{1+\beta}\frac{\bar{k}}{l}. \tag{6.26}$$

Using (6.26), the steady state version of (6.25) can be used to solve for the steady state capital intensity

$$\bar{k} = \left[\frac{\beta(1-\alpha)A}{1+\beta}\right]^{\frac{1}{1-\alpha}}. \tag{6.27}$$

Note that (6.27) gives the same steady state capital intensity as in the one sector model. Thus, the two new forces in the two-sector setting, that affect capital accumulation in (6.25), exactly offset each other to generate the same steady state k as in the one sector case.

An important implication of this result is that countries with different quantities of effective land per person ($\tilde{A}^{1/\alpha}l$) have the same long-run values for k and w. In a two sector model with different goods, w is the purchasing power of wages with respect to manufacturing goods. The purchasing of wages with respect to agricultural goods is w/p. From (6.21), the greater is the quantity of effective land per person, the lower is p and the higher is the purchasing power of the wage in terms of agricultural goods. Thus, there is no lasting "natural resource curse" in this model, once growth is started. Countries with greater natural resources per person will be better off in the long-run.

6.3.6 Transitional Dynamics

We now study the transitional dynamics of the two sector model. Analysis of the two sector model is more challenging than in the one sector model, where plots of the single transition equation are sufficient to determine the model's dynamic properties. One needs a solid background in undergraduate linear algebra to follow the steps. In addition, we also have the problem of the missing initial condition for the price of land that was not an issue in the one sector case.

There is a nice mathematical result that allows us to examine the dynamics of the nonlinear system of equations given by (6.24) and (6.25) in the neighborhood of its steady state by studying a linear system of deviations in k and p^L away from their steady state values. In the neighborhood of the steady state, the nonlinear system and the linear system behave similarly (see, for example, Azariadis (1993 6.3)). While this approach does not tell us about the global dynamics of the system away from the steady state, it does offer clues that may help in a global analysis of the nonlinear system based on numerical methods. See the Technical Appendix for an example of how this approach works in a less difficult setting.

The linear system is formed by taking a linear approximation of the nonlinear system in the neighborhood of the steady state. A linear approximation of a nonlinear equation $f(x) = 0$, where x is a vector of variables and $\bar{x}$ is a vector of particular values of x around which the approximation is constructed, takes the form $f(x) \approx f(\bar{x}) + Df(\bar{x})(x - \bar{x})$. The notation $Df(\bar{x})$ represents a vector of derivatives of the function f with respect to each of its arguments, all evaluated at the point $\bar{x}$. Note that if $f(x) = 0$ is satisfied at $\bar{x}$, then $f(\bar{x}) + Df(\bar{x})(x - \bar{x}) = Df(\bar{x})(x - \bar{x})$.

To take this approach in our setting, begin by writing (6.24) and (6.25) as the following nonlinear system of equations

$$F\left(p_{t+1}^L, p_t^L, k_{t+1}, k_t\right) \equiv k_{t+1} - \left(\frac{1}{1 - \frac{(1-\chi)(1-\alpha)}{1+\beta}}\right)\left[\frac{\beta A(1-\alpha)}{1+\beta}k_t^\alpha - p_t^L l\right] = 0$$

$$G\left(p_{t+1}^L, p_t^L, k_{t+1}, k_t\right) \equiv p_{t+1}^L - \left(1 + \alpha A k_{t+1}^{\alpha-1} - \delta\right)p_t^L + \frac{\alpha(1-\chi)(1-\alpha)A}{(1+\beta)l}k_{t+1}^\alpha = 0.$$

Next take linear approximations of F and G in the neighborhood of the steady state. Then solve the linear system and simplify, using $r = \alpha A\bar{k}^{\alpha-1}$, to get the following linear system

$$\widehat{k}_{t+1} = \left(\frac{\beta(1-\alpha)r}{1+\beta-(1-\chi)(1-\alpha)}\right)\widehat{k}_t - \left(\frac{(1+\beta)l}{1+\beta-(1-\chi)(1-\alpha)}\right)\widehat{p}_t^L \tag{6.28a}$$

$$\widehat{p}_{t+1}^L = -\left(\frac{(1-\chi)\beta(1-\alpha)^2 r^2}{l(1+\beta)(1+\beta-(1-\chi)(1-\alpha))}\right)\widehat{k}_t + \left(1 + \frac{r(1+\beta)}{1+\beta-(1-\chi)(1-\alpha)}\right)\widehat{p}_t^L, \tag{6.28b}$$

where the “^” denotes a deviation of the variable from its steady state value.

The mathematics of linear difference equations system is well developed. To use the available mathematical results, it is convenient to express (6.28) in matrix form as

$$\begin{bmatrix}\widehat{k}_{t+1}\\ \widehat{p}_{t+1}^L\end{bmatrix} = J\begin{bmatrix}\widehat{k}_t\\ \widehat{p}_t^L\end{bmatrix}, \tag{6.29}$$

where

$$J \equiv \begin{bmatrix}\frac{\beta(1-\alpha)r}{1+\beta-(1-\chi)(1-\alpha)} & -\frac{(1+\beta)l}{1+\beta-(1-\chi)(1-\alpha)}\\ -\left(\frac{(1-\chi)(1-\alpha)r}{l(1+\beta)}\right)\left(\frac{\beta(1-\alpha)r}{1+\beta-(1-\chi)(1-\alpha)}\right) & 1 + \frac{r(1+\beta)}{1+\beta-(1-\chi)(1-\alpha)}\end{bmatrix}.$$

The solution to (6.29) can be written as

$$\widehat{k}_t = C_1\lambda_1^t + C_2\lambda_2^t \tag{6.30a}$$

$$\widehat{p}_t^L = C_1 e_{12} \lambda_1^t + C_2 e_{22} \lambda_2^t, \tag{6.30b}$$

where λ_1 and λ_2 are the eigenvalues of J, $\begin{bmatrix} 1 \\ e_{12} \end{bmatrix}$ and $\begin{bmatrix} 1 \\ e_{22} \end{bmatrix}$ are associated eigenvectors, and C_1 and C_2 are undetermined constants (Azariadis (1993, 4.1)).[3]

Note that the dynamic properties of the system are determined by the eigenvalues. If the eigenvalues are less than one, then the deviations from the steady state values will go to zero over time, i.e. the capital-labor ratio and the price of land will converge to their steady state values. If the eigenvalues are greater than one, then the values of the capital-labor ratio and price of land will diverge away from their steady state values.

As mentioned previously, the initial condition for the capital-labor ratio can be used to determine one of the unknown constants, but there is no similar condition for the initial price of land. One way of determining the second constant is to require that the system converge to the steady state. For this approach to work, the parameter values of the model must be such that one of the eigenvalues is less than one and one of the eigenvalues is greater than one. In this case, convergence to the steady state requires that the undetermined coefficient associated with the unstable eigenvalue be set to zero. Setting the coefficient to zero allows one to determine an initial value for the price of land that is associated with a stable path to the steady state.

We can investigate further by using additional results from linear algebra that relate the eigenvalues to J. The sum of the eigenvalues is equal to the *trace* of J and the product of the eigenvalues is equal to the *determinant* of J. These results, and using the fact that $r = \alpha(1 + \beta)/(1 - \alpha)\beta$, gives us two equations to solve for the eigenvalues

$$\lambda_1 + \lambda_2 = 1 + \kappa[1 + \beta + \beta(1 - \alpha)] \tag{6.31a}$$

$$\lambda_1 \lambda_2 = \kappa(\beta + \alpha), \tag{6.31b}$$

where $\kappa \equiv r/[1 + \beta - (1 - \chi)(1 - \alpha)]$. Using (6.31), we can compute the eigenvalues for different parameter settings.

Let's look at the case where $\alpha = 1/3, \chi = 2/3$, and $\beta = 1/2$. In this case we find that $\lambda_1 = 0.3490$ and $\lambda_2 = 2.8031$. So, we have one stable eigenvalue and one unstable eigenvalue. Convergence to the steady state requires that the coefficient associated with the unstable eigenvalue be set to zero, i.e. $C_2 = 0$. Furthermore, the definition of an eigenvector (Azariadiz (1993, pp.123–127)) gives us

$$J \begin{bmatrix} 1 \\ e_{12} \end{bmatrix} = \lambda_1 \begin{bmatrix} 1 \\ e_{12} \end{bmatrix},$$

[3]To review eigenvalues and eigenvectors consult the technical appendices of Azariadis (1993) or an undergraduate linear algebra textbook.

which implies $e_{12} = \frac{\lambda_1 - J_{11}}{J_{12}} = \frac{\omega - \lambda_1}{\frac{l(1+\beta)}{1+\beta-(1-\chi)(1-\alpha)}}$, where $\omega \equiv \frac{\beta(1-\alpha)r}{1+\beta-(1-\chi)(1-\alpha)}$. The remaining unknown is the coefficient associated with the stable eigenvector, which is determined by the initial condition for the capital-labor ratio. Applying the initial condition to (6.30a) when $t = 0$, yields $\widehat{k}_0 = k_0 - \bar{k} = C_1$.

Our complete solution is then

$$k_t = \bar{k} + \left(k_0 - \bar{k}\right)\lambda_1^t \tag{6.32a}$$

$$p_t^L = \bar{p}^L + \left(k_0 - \bar{k}\right)\left[\frac{\omega - \lambda_1}{\frac{l(1+\beta)}{1+\beta-(1-\chi)(1-\alpha)}}\right]\lambda_1^t. \tag{6.32b}$$

Note, for the parameters chosen above we have $\omega = 0.3913$, so the expression in the squared bracket of (6.32b) is positive. If we start with a capital-labor ratio below the steady state, the capital-labor ratio rises over time because the negative expression, $\left(k_0 - \bar{k}\right)\lambda_1^t$, becomes smaller over time. The price of land will start below its steady state value, because the second expression on the right-hand-side of (6.32b) is negative when $k_0 < \bar{k}$ and falls in absolute value over time. Land prices rise over time creating capital gains for land owners. However, as our previous discussion of (6.24) suggests, the positive *growth rate* of land prices is falling over time, causing the return to land to fall with the return to physical capital.

Remember that the analysis we just conducted is only valid when the economy is close to its steady state. To conduct a global analysis, numerical methods can be used to compute the dynamic path (e.g. see Hansen and Prescott (2002)). If the parameter values are similar to the case we just examined, then it is likely that for an initial value of k there is a unique initial value for p^L that is on the unique convergent path to the steady state. One can guess a value for the initial land price and then use (6.25) to compute next period's capital-labor ratio and (6.24) to compute next period's land price. Iterating in this manner, the entire dynamic path can be solved. If the path is not converging to the steady state, then an alternative initial guess for p^L is tried. Searching over a range of initial values for p^L will eventually identify the convergent equilibrium path.

6.3.7 International Trade

We now examine the dynamics of the economy when there is international trading of goods across countries. We assume that the economy is "small" in the sense that its behavior cannot affect the international or worldwide relative price of agricultural goods, p^w. This means that the relative price of food is no longer determined endogenously by domestic demand and supply, i.e. (6.21) no longer applies. Instead p^w is taken as an exogenous variable that is determined outside the model by demand and supply conditions in the worldwide market for goods.

The first important implication of opening the economy to trade is that one of the domestic forces that caused the fraction of the work force in the traditional sector to remain constant, as the economy accumulates capital, is now absent. Recall that capital accumulation raises the marginal product of labor and wages in the modern sector, which reduces the demand for labor in the traditional sector. In a closed economy, higher wages increases demand for all goods and services and increases the relative price of the agricultural good. This latter effect raises the demand for labor in the traditional sector. Under our assumptions, these two effects exactly offset and the fraction of labor employed in the traditional sector remains constant. However, the latter effect is now absent because the relative price of agricultural goods is fixed by market conditions at the world level, not the domestic level. An increase in the demand for agricultural goods simply results in fewer exports, or more imports, at the same world price.

In an open economy the fraction of the workforce in the traditional sector is given by

$$1-\pi_t=\left(p^w\tilde{a}\right)^{1/\alpha}\frac{l_t}{k_t}. \tag{6.33}$$

So, capital accumulation serves only to raise the cost of labor in the traditional sector, causing that sector to contract with development.

Furthermore, we see that if the world price for agricultural goods increases, then the traditional sector will expand. This result is related to concerns that if a developing economy begins to trade internationally it may lead to "de-industrialization" (see, for example, Williamson (2011)). The idea is that developing countries are likely to have a comparative advantage in producing agricultural goods. This implies that their closed economy relative price of agricultural goods is below the world relative price. When they open to trade, the relative price will rise up to p^w. The higher the value of p^w, the more labor will be reallocated away from the modern sector and toward the traditional sector.

Many economists believe that they are several connections between the relative size of the modern sector and economic growth in the economy as a whole. We will examine the consequence of "de-industrialization" for growth and welfare in developing countries both here and in the models that we encounter in Chaps. 7 and 8.

With an exogenous price of agricultural goods, the dynamics of the economy change. With international trade of goods, Eqs. (6.24) and (6.25) become

$$p_{t+1}^L=\left(1+\alpha A k_{t+1}^{\alpha-1}\right)p_t^L-\left(\tilde{a}p^w\right)^{1/\alpha}\alpha A k_{t+1}^{\alpha-1} \tag{6.24$'$}$$

and

$$k_{t+1}=\left(\frac{1}{1-\frac{l\left(\tilde{a}p^w\right)^{1/\alpha}}{k_{t+1}}}\right)\left[\frac{\beta A(1-\alpha)}{1+\beta}k_t^\alpha-p_t^L l\right]. \tag{6.25$'$}$$

The steady state equilibrium is

$$\bar{p}^L = \left(\tilde{a}p^w\right)^{1/\alpha}. \tag{6.26'}$$

$$\bar{k} = \left[\frac{\beta(1-\alpha)A}{1+\beta}\right]^{\frac{1}{1-\alpha}}. \tag{6.27'}$$

Comparing to the steady state in the closed economy, we see that the price of land will differ but not the capital-labor ratio. This is due to two opposing effects of a higher relative price of agricultural goods on capital accumulation. First, a higher value of p^w raises land prices (see (6.26′)). Higher land prices reduce capital accumulation by diverting more saving to land purchases (see (6.25′)). Second, a higher value of p^w increases the employment share in the traditional sector (see (6.33)). A smaller employment share in the modern sector reduces the crowding of capital and raises the capital-labor ratio. In the current model, these two opposing effects just cancel. So de-industrialization only applies to the fraction of the workforce in the traditional sector. There is no negative long-run consequence for the capital-labor ratio in the modern sector.

If the long-run capital-labor ratio remains unaffected by opening the economy to trade, then the wage in terms of manufacturing goods will also be unchanged. However, if opening the economy is associated with a higher price of agricultural goods, then the purchasing power of wages with respect to agricultural goods must fall.

The same argument applies to the return to capital. Measured in units of the manufacturing good it remains the same, but it falls in units of agricultural goods. Finally, the long-run rental rate on land, $r^L = \left(\tilde{a}p^w\right)^{1/\alpha}\alpha A\bar{k}^{\alpha-1}$, is now positive and increasing in the world price of agricultural goods. It is higher both in units of the manufacturing good and in units of the agricultural good (since $1/\alpha > 1$).

The steady state welfare effects associated with international trade can be determined by writing out the indirect utility function of the steady state household as $(1+\beta)\ln w + \beta \ln R - (1-\chi)\ln p^w$. In the steady state w and R are unaffected by opening the economy, while the price of food rises. Thus, the developing country is worse off in the long-run. This represents a counterexample to the notion that welfare must rise as a result of trade.

6.3.8 Transitional Dynamics in an Open Economy

We can study the transitional dynamics of the open economy in the same manner as we did for the closed economy. The linear approximation to the dynamic system in the open economy is

Table 6.1 Open-economy dynamics

π	$\lambda_1 = \alpha/\pi = 1/3\pi$
0.5	0.6667
0.7	0.4761
0.9	0.3703

$$\begin{bmatrix} \widehat{k}_{t+1} \\ \widehat{p}^L_{t+1} \end{bmatrix} = P \begin{bmatrix} \widehat{k}_t \\ \widehat{p}^L_t \end{bmatrix}, \tag{6.34}$$

where

$$P \equiv \begin{bmatrix} \dfrac{\alpha}{\pi} & -\dfrac{(1-\pi)\bar{k}}{\pi\bar{p}^L} \\ 0 & 1+r \end{bmatrix}.$$

One can show that there is at most one stable eigenvalue associated with the linear system (see *Problem 13*). So if there is a convergent path to the steady state, it is unique. For the same parameter values used in the closed economy setting, there is an array of possibilities for the open economy that depend on the international price of agricultural goods and the corresponding size of the modern sector. The lower the price of agricultural goods, the greater the size of the modern sector. Table 6.1 below exhibits some possibilities and the corresponding value of the stable eigenvalue.

Note that the absolute value of the stable eigenvalue can be higher or lower than in the closed economy case. However, for the stable eigenvalue to be less than in the closed economy π would have to be very close to 1. This means that the open economy is likely to converge more slowly to the steady state than in the closed economy. For the parameter values chosen previously, the steady state value of π in the closed economy case is a little over 0.85 and $\lambda_1 = 0.349$. If opening the economy raises the relative price of food, then the value of π would be lower than 0.85. From Table 6.1 we see the stable eigenvalue will be greater than in the closed economy case and convergence to the steady state will be slower.

6.4 Declining Budget Shares Spent on Food

One feature of development, not yet captured by the model, is the declining shares of household budgets devoted to food. Over two centuries of modern growth, from the mid-eighteenth to the mid-twentieth century, the household budget share devoted to food fell from about 60% to about 30% (Abdus and Rangazas 2011, Table 5). In contrast, if one interprets the traditional sector good as food, the current version of the model predicts a constant budget share.[4]

[4]This need not be the only interpretation because primary products such as wood and coal might also be included.

One way of generating a declining budget share is to change preferences so that a minimum amount of food consumption is necessary for survival—what is known as a subsistence constraint. We can incorporate a subsistence constraint by writing the utility benefits from food consumption as $(1-\chi)\ln\left(o_{1t}-\bar{o}\right)$, where $\bar{o}$ is the minimum amount of food consumption needed to survive. With this new specification, the household's first period budget shares become

$$\frac{y_{1t}}{w_t}=\frac{\chi}{1+\beta}\left(1-\frac{p_t\bar{o}}{w_t}\right) \tag{6.35a}$$

$$\frac{p_t o_{1t}}{w_t}=\frac{1-\chi}{1+\beta}+\frac{\beta+\chi}{1+\beta}\frac{p_t\bar{o}}{w_t}. \tag{6.35b}$$

As wages rise relative to food prices, i.e. as the purchasing power of wages in terms of food rises, the share devoted to nonfood rises and the share devoted to food falls.

Fogel (1994, p.377) points out that the concept of a fixed level of subsistence consumption is potentially misleading. He suggests that are there are numerous possible combinations of consumption and health that are consistent with subsistence in a low income setting. Abdus and Rangazas (2011) follow Fogel's lead and provide an alternative approach to subsistence consumption that relates food consumption directly to health. They use the approach to explain why food consumption, measured by caloric in-take, has remained relatively constant over the course of development. This, in turn, explains the falling budget share devoted to food as income rises.

We can change preferences to include health benefits by defining u_{1t} as $u_{1t}=\chi\ln(y_{1t}Q_t)+(1-\chi)(\ln o_{1t}Q_t)$, where the utility benefits of consumption are affected by a health index Q_t. The more healthy the individual, the more he enjoys the activities that give him pleasure. The health index is defined as $Q_t=Q_t^*e^{-\kappa(b_t-b^*)^2/2}$. The variable Q_t^* is an index of exogenous factors that affect health such as air quality, public sanitation, public health services, and working conditions. The endogenous component of the health index relates health to body weight. Body weight is a well-documented factor in determining health and longevity.

A standard way of recording body weight is the Body Mass Index (BMI), defined as weight in kilograms divided by the square of height in meters. The BMI that maximizes health is in the range of 25–26 (Fogel 1994 and Costa 1996). Being too light or too heavy compromises a person's health. In the health index term, the actual BMI is denoted by b_t and the optimal BMI by b_t^*. In this section the all-purpose parameter κ quantifies the negative health effects of allowing body mass to deviate from its optimal level.

We relate food and work to body weight using the science of energy balance for the human body (Whitney and Cataldo 1983 and Fogel 1994). Increased calories consumed as food will raise an individual's body weight unless an equal amount of energy is expended in (i) basal metabolism (the energy cost associated with keeping the body alive at rest), (ii) the thermic effect of food consumption and digestion (about 10% of the calories consumed) or (iii) physical activity. Calories consumed in excess of calories spent are stored in the body as fat. For each 3500 kcal taken in excess of expenditures, one pound of body fat is stored.

Scientists have estimated an energy balance relationship that relates daily caloric in-take and calories expended in physical activity to the daily change in body weight. We are not interested in daily changes in body weight but rather in the body weight that an individual can maintain in the long term given a regular regiment of daily caloric consumption and physical activity. Toward this end, we use a steady state version of the energy balance equation for daily changes in body weight. We solve the difference equation that determines the daily change in body weight for the steady state or sustainable body weight associated with different levels of caloric intake and energy expenditure that are maintained over the years. Furthermore, we express the relationship in terms of the body mass index.

The steady state energy balance equation for an individual who reaches adulthood in period-t is given by

$$b_t = \frac{0.9o_{1t} - \bar{b}_1 - \varepsilon_t}{\bar{b}_2 T_t^2} \tag{6.36}$$

where $\bar{b}_1$ and $\bar{b}_2$ are positive parameters that are determined by the physics of energy balance, ε_t measures the required average daily calories expended per unit of work, and T_t is adult height.[5] The parameters $\bar{b}_1$ and $\bar{b}_2$ capture the linear effect of bodyweight on basal metabolism and the energy expended in basic survival activity. The variables ε_t and T_t are both treated as exogenous here, i.e. we do not model the choice of occupation and the many factors that cause people to become taller across generations.[6]

The first order condition associated with choosing the optimal demand for the agricultural good now becomes

$$\frac{1-\chi}{\chi}\frac{1}{o_{1t}} + \frac{(b^* - b_t)0.9}{\bar{b}_2 T_t^2}\frac{\kappa}{\chi} = \frac{p_t}{y_{1t}}. \tag{6.37}$$

Equation (6.37) equates the marginal utility benefit of calories consumed (left-hand-side) to the marginal utility cost of purchasing calories (right-hand-side). If b_t is less than b_t^*, then there are two benefits to consuming calories: the direct pleasure of food

[5]For further discussion and application of the energy balance relationship see Abdus (2007).

[6]Establishing an endogenous intergenerational connection between child nutrition and adult height is difficult. In the nineteenth century, adult height in England showed no trend, despite a rise in per capita income. This puzzling observation has given rise to a literature that identifies a variety of factors that potentially affect adult height (Kirby (1995), Komlos (1998), and Voth and Leunig (1996)). Modern studies also reveal the complex determination of adult height. For example, food intake is not strongly correlated with height (Graham et al. (1981), Ashworth and Millward (1986), and Mitchell (1962)).

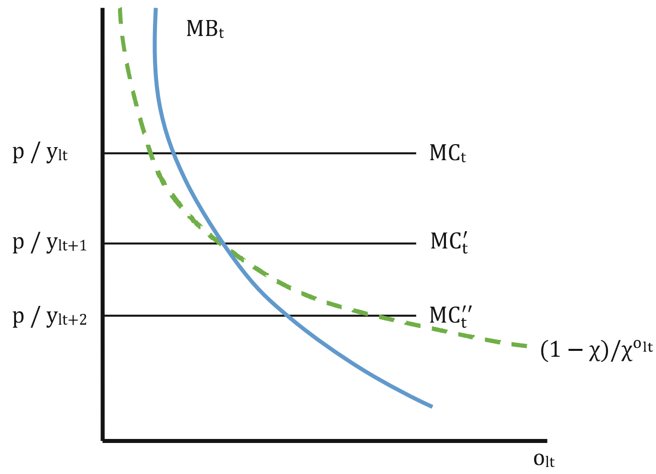


Fig. 6.1 Optimal food consumption

consumption (first term) and the increase in body mass that makes the individual healthier (second term).

Sketching the marginal benefit and marginal cost of food (other things constant) produces Fig. 6.1. The dashed curve gives the marginal benefit of food intake when there are no health benefits ($\kappa = 0$). Along the dashed curve, y_{1t} and o_{1t} increase proportionally with wages as the economy grows and the value of the forgone marginal consumption falls, for a constant p. A constant budget share devoted to food.

With $\kappa > 0$, the marginal benefit of food exceeds the pleasure of eating when incomes are low, causing food to be a larger fraction of the budget than when $\kappa = 0$. As income and y_{1t} rise over time, the opportunity cost of consuming calories falls and o_{1t} *rises less than proportionally* along the marginal benefit curve, resulting in a falling fraction of the budget devoted to food.

If the energy requirements of work fall with development, and individuals respond by expending less energy, b_t increases for a given level of food consumption. This causes the marginal benefit curve to shift left over time and the optimal food consumption may remain constant or fall even as income grows. This gives a second reason for a declining budget share devoted to food.

Food consumption is most likely to rise when the health effects of food are small, i.e. when b_t is close to b_t^*. In this case the marginal benefit curve is close to the dashed-curve (where y_{1t} and o_{1t} increase proportionally). This helps explain the rise in food consumption in the late twentieth century when body mass was close to optimal.

6.4.1 A Demand-Side Explanation for the Structural Transformation

The subsistence constraint also provides another mechanism that helps to explain the structural transformation. Recall that in the closed economy version of the two sector model, the share of employment in the traditional sector is constant (see (6.22)). With a subsistence constraint this will not be the case. If there is economic growth that raises the purchasing power of wages across both goods, the fraction of employment in agriculture falls.

To see the reason for this result, let's go back to the simple subsistence constraint specification that does not introduce health benefits directly. With the subsistence constraint assumption, one can use the equilibrium condition in the market for food to write a new version of (6.20)

$$\frac{l}{k_t}(p_t\tilde{a})^{1/\alpha} = 1 - \pi_t = \frac{1-\chi}{1+\beta}(1-\alpha) + \frac{\beta+\chi}{1+\beta}\frac{\bar{o}}{w_t/p_t}(1-\alpha). \tag{6.38}$$

If $\bar{o} = 0$, then we are back to (6.22), where the labor shares across sectors are fixed. With subsistence consumption, if real wages grow in units of the traditional good, then the traditional sector employment share will shrink.

For example, for a given k_t, a rise in $\tilde{A}$ in our previous analysis caused a proportional decline in p_t, so that the labor shares remained constant. Intuitively, the rise in $\tilde{A}$ caused a proportional increase in the supply *and the demand* for agricultural goods, as the real wage in traditional good units rose, that kept the labor employed in the traditional sector constant. Now, a rise in $\tilde{A}$ will raise supply more than demand because the subsistence component of demand is not a function of real income. The amount of labor needed in the traditional sector to meet demand falls, causing a fall in the traditional sector's labor share. In summary, a higher value of $\tilde{A}$ makes it easier to satisfy the basic subsistence component of food consumption and releases labor to work in the modern sector.[7]

6.5 Conclusion

Sustained modern growth began around 1800. We think of the beginning of modern growth as coinciding with the appearance of factory-based production where man-made physical capital first became an important input. Ideas about what equipment and plants to build, and how to power and operate them using labor, had to evolve until the state of "technology" reached a threshold where these factories became sufficiently productive. The factories had to be productive enough

[7] In a complete general equilibrium analysis, there would also be a reduction in k as workers move into the modern sector for a given aggregate capital stock. For an analysis of the structural transformation that depends heavily on the demand side effects stemming from subsistence constraints see Herrendorf et al. (2014).

to pay workers what they earned on the farm and capital investors a return at least as high as they could receive by owning land.

Country-specific characteristics caused the threshold level of technology to differ across the world, affecting the timing of the take-off toward sustained growth. The required threshold level of technology is greater (i) the more abundant and productive is the country's land, (ii) the weaker are the country's institutions in creating the property rights and markets associated with producing and selling at a large and impersonal scale, and (iii) the stronger are the vested interests in the country that block the use of new technologies.

In addition, countries may differ in the fundamentals that determine the take-off of mass education. Preventing child labor and encouraging schooling will create sustained gains in human capital and worker productivity that complements the growth in physical capital. A delay in the timing of the modern growth take-off, by inhibiting the growth in physical and human capital, is one reason that countries fall behind, leading to the large income gaps we see across the world today.

Another source of today's income gaps is the pace of the structural transformation as an economy goes through the lengthy process of shifting resources away from traditional means of production to modern capital-intensive production. In a two-sector setting, the accumulation of capital is more complicated than is suggested by one sector models. First, the movement of labor out of the traditional sector crowds the stock of physical capital and lowers the capital-labor ratio (even if there is no population growth). Second, the presence of land provides a substitute asset through which households can save. The accumulation of capital may, in turn, exert a causal feed-back effect on the migration of labor and the price of land.

In a closed economy with two distinct goods, agricultural and industrial, we saw that the pace of the migration of labor across sectors is not affected by the accumulation of capital. More capital raises the marginal product of labor and wages in the modern sector, increasing the migration of labor toward the modern sector. However, higher income also raises the price of food, which slows the migration of labor away from the traditional agricultural sector. In our model, these two effects exactly offset and the share of labor in the modern sector is independent of capital accumulation. Thus, capital accumulation alone may not be sufficient to increase the modern sector's labor share.

Opening the economy implies the relative price of agricultural goods is determined outside of the country's domestic economy. This removes the positive effect of capital accumulation on agricultural prices, implying a migration of labor out of the traditional sector as the capital stock increases. For this reason, internationally trading goods allows capital accumulation to pull labor away from the traditional sector. This result provides a possible counterexample to the De-industrialization hypothesis. The De-industrialization hypothesis argues that opening the economy to the trade of goods slows the structural transformation in countries that experience a rise in the relative price of agricultural goods. Here we see that a jump in the relative price of agricultural goods does cause a decrease in the modern sector labor share, but it can also prevent further increases in agricultural prices as the economy grows via capital accumulation. This later effect serves to speed the structural transformation. This is just the beginning of our analysis of the effect of trade on growth and the De-industrialization hypothesis. We re-visit this topic in Chaps. 7 and 8.

6.6 Exercises

6.6.1 Questions

1. Describe economic growth and cross-country income inequality before and after 1800.
2. Explain the difference between the following concepts.
   (a) price of land
   (b) rental rate for land
   (c) return to owning land
3. In the traditional economy, what happens to the following variables if the economy experiences a one-time unanticipated improvement in TFP ($\tilde{A}$)? A one-time increase in population size (N)?
   (a) wage paid to a worker
   (b) price of land
   (c) rental rate for land
   (d) return to owning land
4. Describe how we established the conditions required for modern growth and industrialization to begin.
5. What are some of the potential causes of the advance of technology in the pre-modern period?
6. Discuss some of the features of an economy that might delay the onset of modern economic growth.
7. Suppose that roads are built that causes both A and $\tilde{A}$ to double. How does this change the likelihood of seeing a modern sector appear, if at all? Explain.
8. Is population growth enough to shrink the relative size of the traditional sector and expand the relative size of the modern sector? Explain.
9. Discuss the major differences between capital accumulation in a two sector versus a one sector economy.
10. What is an asset bubble? Is an asset bubble possible in our two sector model?
11. Explain the role of schooling in initiating or accelerating modern growth.
12. Is there a "natural resources curse" in the two sector model? In other words, are households in an economy with a higher stock of effective land worse off?
13. Explain why when the economy is open to the trade of goods, the traditional sector shrinks with capital accumulation, whereas in a closed economy the size of the traditional sector was constant.
14. How might international trade cause "de-industrialization"? Does this lower steady state welfare in the two sector model?
15. Do we know that opening the economy to trade lowers the welfare of a developing country in transition to the steady state?

16. Explain why taking account of the health benefits of food helps explain why the budget share devoted to the agricultural good declines with development.
17. Many economists argue that an increase in agricultural productivity shrinks the employment share of the traditional sector. Does this happen in our models?
18. Use the models of this chapter to explain the growth facts G8, G9, and G10.

6.6.2 Problems

1. Suppose the current price of land is \$100 per acre, a value that is expected to stay constant over time ($p_t^L = p_{t+1}^L =$ \$100). Assuming that one can rent land for \$10 each period ($r_t^L = r_{t+1}^L =$ \$10), what is the return to land? What is the rate of return? Redo your calculation if the price of land was \$50. Under which scenario is it less likely that modern firms will operate profitably? How is this effect included in the threshold condition?
2. Why can $\beta(1 - \alpha)/(1 + \beta)$ be interpreted as the rate of saving out of income?
3. Assume $\alpha = 1/3$, $A = 1$, and $\delta = 0.10$. Compute the threshold value that D_t must exceed to introduce profitable firms and modern growth in each of the following scenarios.
   (a) $\tilde{A} = 1, \tilde{l} = 1, \beta(1-\alpha)/(1+\beta) = 0.20$
   (b) $\tilde{A} = 2, \tilde{l} = 1, \beta(1-\alpha)/(1+\beta) = 0.20$
   (c) $\tilde{A} = 1, \tilde{l} = 2, \beta(1-\alpha)/(1+\beta) = 0.20$
   (d) $\tilde{A} = 1, \tilde{l} = 1, \beta(1-\alpha)/(1+\beta) = 0.10$
4. Suppose modern technology is growing at 1% per year and each period in the model lasts 20 years. In other words, we assume the following growth equation for the technology index: $D_t = (1 + d)^t$, where $d = (1.01)^{20} - 1 = 0.2202$. Take the natural log of the expression D_t and plot it as a function of time, i.e. plot ln $D_t = \ln(1 + d) \times t \approx dt$ (because $\ln(1 + d) \approx d$ when d is a small value). If the threshold value of D_t is 200, at what value of t will the country see modern production? If technology goes more slowly, at 0.5% annually, how long will it take for the modern sector to appear?
5. *Scrimmage for section 1*. Provide a derivation and an economic interpretation for the following equations: (6.3), (6.4), and (6.8).
6. Sketch the relationship between population size and wages based on (6.5a). How would an increase in $\tilde{A}$ affect wages? Remember Malthus conjectured that an increase in $\tilde{A}$ would NOT have a lasting effect on wages. Why did he did he believe this?
7. *Scrimmage for section 2*. Provide a derivation and an economic interpretation for the following equations: (6.10) and (6.11).
8. Suppose $D_t = \tilde{D}_t$. Use (6.9a), and the marginal product of labor expressions in each sector, to sketch the demand for labor in each sector with the common wage on the vertical axis and the fraction of the work force in the modern sector on the

horizontal axis. This is how the static Specific Factors model is typically presented. Use the diagram to show how changes in $L, \tilde{A}, A$, and K_t affect w_t and π_t.

9. *Scrimmage for section 3*. Provide a derivation and an economic interpretation for the following equations: (6.19), (6.25), and (6.28), (6.35).
10. How does the diagram in *Problem* 8 change when you have a closed economy where two distinct goods are traded at the relative price p_t? Write out the value of the marginal product expressions in units of the manufacturing good in the same manner as in *Problem* 8 and use (6.21) to eliminate the relative price in terms of its closed economy determinants. It will make it more clear that there is a unique equilibrium if the term $\pi_t N_t$ is eliminated from both wage equations before making the sketch. Use the diagram to show how changes in $L, \tilde{A}, A$, and K_t affect w_t and π_t. Next assume that the economy is perfectly open and repeat the exercise. In addition, show how a rise in the international determined relative price affects the labor share.
11. Fun with eigenvalues and dynamics. Check the robustness of a unique and convergent path for the closed economy described by (6.24) and (6.25). Keep the capital share at 1/3 and consider combinations of values for χ and β that are between 0 and 1. Report your findings in a table.
12. Derive the indirect utility function of a household in the open economy version of the two sector model. Use the indirect utility function to argue that steady state welfare falls when developing countries with a comparative advantage in agricultural goods open their economies to trade.
13. Without using numerical values for parameters, argue that there at most one stable eigenvalue to the dynamic system given by (6.34).
14. *Scrimmage for section 4*. Provide a derivation and an economic interpretation for the following equations: (6.35), (6.37), and (6.38).
15. Use (6.37) to explain the configuration of Fig. 6.1.

References

Abdus S (2007) Obesity: A dynamic structural model of individuals' body weight choice. Cornell University, Ithaca, New York.

Abdus S, Rangazas P (2011) Adult nutrition and growth. Review of economic dynamics 14 (4):636–649.

Acemoglu D, Robinson J (2012) Why nations fail: The origins of power, prosperity, and poverty. Crown Publishers, New York.

Ashworth A, Millward D (1986) Catch-up growth in children. Nutrition reviews 44(5):157–163.

Azariadis C (1993) Intertemporal macroeconomics. Blackwell, Oxford.

Clark G (2007) A farewell to alms. Princeton University Press, Princeton, New Jersey.

Costa D (1996) Health and labor force participation of older men, 1900–1991. Journal of economic history 56(1):62–89.

Crafts N (1995) Exogenous of endogenous growth? The industrial revolution reconsidered. Journal of economic history 55(4):745–772.

Fogel R (1994) Economic growth, population theory, and physiology: The bearing of long-term processes on the making of economic policy. American economic review 84(3):369–395.

Galor O (2005) From stagnation to growth: A unified theory. In: Aghion P, Durlauf S (eds) Handbook of economic growth. Elsevier North-Holland, Amsterdam.

Galor O (2011) Unified growth theory. Princeton University Press, Princeton, New Jersey.

Galor O, Weil D (2000) Population, technology, and growth: From Malthusian stagnation to modern economic growth. American economic review 90(4):806–828.

Graham GG, Creed HM, MacLean Jr WC, Kallman CH, Rabold J, Mellits ED (1981) Determinants of growth among poor children: Nutrient intake –achieved growth relationships. The American journal of clinical nutrition 34:539–544.

Habbakuk J (1962) American and British technology in the Nineteenth century. Cambridge University Press, Cambridge, UK.

Hansen G, Prescott E (2002) Malthus to Solow. American economic review 92(4):1205–1217.

Herrendorf B, Rogerson R, Valentinyi A (2014) Growth and the structural transformation. In: Aghion P, Durlauf S (eds) Handbook of economic growth, vol. 2. Elsevier, Amsterdam.

Kirbey P (1995) Causes of Short stature among coal-mining children 1823–1850. Economic history review, 687–699.

Komlos J (1998) Shrinking in a growing economy? The mystery of physical stature during the industrial revolution. Journal of economic history 58(3):779–802.

Lucas R (2002) Lecture on economic growth. Harvard University Press, Cambridge, Massachusetts.

Maddison A (1995) Monitoring the world economy, 1820–199. Organisation for Economic Co-operation and Development, Paris.

Mitchell H (1962) Nutrition in relation to stature. Journal of the American dietetic association 40:521–524.

Mourmouras A, Rangazas P (2009) Reconciling Kuznets and Habbakuk in a unified growth model. Journal of economic growth 14(2):149–181.

Parente S, Prescott E (2000) Barriers to riches. MIT Press, Cambridge, Massachusetts.

Voth H, Leunig T (1996) Did smallpox reduce height? Stature and the standard of living in London, 1770–1873. Economic history review 49(3):541–566.

Whitney E, Cataldo C (1983) Understanding normal and clinical nutrition. West-Wadsworth Publishing Company, Belmont, California.

Williamson J (2011) Trade and poverty: When the third world fell behind. MIT Press, Cambridge, Massachusetts.

Wage and Fertility Gaps in Dual Economies

7

This chapter begins our analysis of two sector models where markets may be missing. Here we focus on the fact that wages are lower and fertility is higher in the traditional sector than in the modern sector of economies. This has important consequences for economic growth for two reasons. First, as suggested by growth fact *G11*, the wage gap indicates the allocation of labor may be inefficient—the movement of labor from the traditional sector to the modern sector should raise average labor productivity. Second, if fertility falls as households move from the traditional sector to the modern sector, then population growth will decline making it easier to increase physical capital per worker.

Sections 7.1, 7.2, 7.3, and 7.4 lay out the basic theory that explains wage and fertility gaps, based on an updated and revised version of the survey in Mourmouras and Rangazas (2013a). Sections 7.5, 7.6, and 7.7 examine the assumptions and interpretation of the theory and provide some new applications.

7.1 Wage and Fertility Gaps

There is substantial evidence documenting the presence of large gaps in worker productivity across agricultural and nonagricultural sectors in the early stages of development (Caselli and Coleman 2001; Gollin et al. 2004; Gollin et al. 2014). To the extent that these productivity gaps reflect gaps in the marginal product of labor, they imply labor is inefficiently allocated across sectors and that Total Factor Productivity (TFP) and aggregate economic growth increase as labor migrates from the low productivity traditional agricultural sector to the high productivity manufacturing sector. Indeed, differences in the allocation of labor across sectors have been shown to explain a significant portion, and in some cases the majority, of TFP differences across countries (Temple and Woessmann 2006; Chanda and Dalgaard 2008; Restuccia et al. 2008; Vollrath 2009a).

There is also evidence that the productivity gaps are more closely connected to differences in production method and institutional arrangements regarding

S. Das et al., *Economic Growth and Development*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-319-89755-4_7

ownership, than to the type of goods produced per se. Even *within urban areas* of developing countries, workers involved in modern firm-based production earn much higher wages than those involved in traditional family-based production (Rosenzweig 1988, pp. 756–757; La Porta and Shleifer 2014). For this reason it may be more accurate to refer to gaps between traditional and modern sectors, rather than between agricultural and nonagricultural sectors. It is primarily because of data limitations that most of the focus is on productivity gaps between farm/rural workers and nonfarm/urban workers.

Table 7.1 gives examples of two dual economies—one from the historical United States and one from contemporary Sub-Saharan Africa. Both economies are dominated by traditional agriculture and have large productivity gaps between nonfarm and farm workers. The sixfold productivity gap in Africa today is particularly large, although it is likely an overestimate due to unmeasured output on family and communal farms that is missed by national income accounting (see, for example, Jerven 2013). The available survey data on wages and consumption reveals much smaller gaps between non-farm and farm or between urban and rural workers in Africa (Mourmouras and Rangazas 2007, pp. 38–39), Henderson (2010, Figure 4), and Vollrath (2009a, 2014).[1] Actual gaps in wages and consumption in current developing countries are much closer to the 2 to 3-fold range witnessed in United States history, but are still larger. For example, Young (2013) finds consumption gaps of 4 for a set of developing countries. We discuss conceptual and measurement issues associated with these gaps in more detail in Sect. 7.7.

Fertility is a related determinant of economic growth that differs significantly across sectors. Table 7.1 indicates that in the early stages of development average fertility is high. Moreover, fertility is much higher in the traditional sector than in the modern sector; 1.5 times higher in both of the dual economies in Table 7.1. This fact suggests that the movement of households from the traditional sector to the modern sector may lower the economy's fertility rate. Reductions in fertility and population growth associated with labor migration to urban areas has the potential to raise labor productivity growth by increasing the accumulation of physical capital per worker and by reducing the fraction of the work force made up of relatively low-productive children and young adults. Thus, migration of labor from the traditional to the modern sector can increase economic growth directly, due to the gap in labor productivity, and indirectly by reducing fertility.

These observations raise the question of why large gaps in productivity and wages persist for decades. As indicated in Table 7.1, the fraction of the economy's work force employed in the relatively low-productivity traditional sector is substantial. Why doesn't labor quickly flow into the more productive and higher-paying modern sector to quickly eliminate the wage gaps?

---

[1]The same measurement issue can be raised about the historical gaps in the U.S., but Alston and Hatton (1991) make adjustments for non-cash payment and cost of living differences and find almost twofold real wage gaps as late as 1940. In fact, Herrendorf and Schoellman (2013) find similar wage gaps in the present day U.S.

Table 7.1 Dual economies

% Modero	Productivity gap	Total fertility rate	Fertility gap
United States—beginning of the nineteenth century			
20	2 to 2.5	7	0.67
Sub-Saharan Africa—beginning of the twenty-first century			
29	6	5	0.66

Notes: United States—*% Modern* is the percentage of labor in non-agriculture in 1800 (David 2005, Table 2.2, column 2), *Productivity Gap* is the ratio of nonfarm to farm productivity in 1840 (David 1967, *Total Fertility Rate* is average fertility rate for the entire economy in 1800 Haines 2000, Table 4, *Fertility Gap* is urban to rural fertility rate in 1800 Greenwood and Seshadri 2002). Sub-Saharan Africa—*% Modern* is the percentage of labor in non-agriculture in 1996 (Temple 2005, Table 7.1), *Productivity Gap* is the ratio of nonfarm to farm productivity in 1996 (Temple 2005, Table 7.1), *Total Fertility Rate* is average fertility rate for the entire economy in 1999 (Galor (2005) and 2010 (UNESCO-UIS Statistics in Brief)), *Fertility Gap* is the ratio of the urban to rural fertility rate between 1996 and 2003 for 24 African countries (computed from Shapiro and Gebreselassie 2008, Table 7.1)

In this chapter we use a dual economy approach to examine different explanations for how large productivity gaps can persist in equilibrium for long periods of time. We relate the theories of productivity gaps to theories of the fertility gap in order to provide a unified explanation of both gaps and their effect on economic development.

We begin in Sect. 7.2 with a survey of the literature that attempts to explain wage and productivity gaps in models assuming complete markets for land and labor. The complete markets approach is useful in emphasizing alternative sources of wage gaps (e.g. migration costs and education differences). However, we argue that these sources are not sufficient to explain the persistence of large wage gaps.

In the main part of the chapter, Sects. 7.3 and 7.4, we discuss explanations for productivity and fertility gaps that assume the absence of factor markets in the traditional sector. More specifically, we review models where (i) the production technology in the traditional sector is family/village owned and operated and (ii) there is an absence of formal markets for at least some of the inputs used in production. Thus, we focus on a fundamental source of dualism, the absence of complete markets, as stressed in the older literature (Lewis 1954; Ranis and Fei 1961).

We think of the existence and effectiveness of markets as differing across the two sectors as the markets for inputs used in the modern sector will generally be more developed, supported to a greater degree by formal legal institutions. In the traditional sector, informal family and tribal institutions govern the allocation of land and labor. Under this interpretation of the dual economy, the disappearance of the traditional sector is associated with the spread of formal markets for productive inputs.[2] In addition to

[2]One can also go deeper and explore the reasons why the formal institutions and related government infrastructure needed to establish well-functioning markets are not adequately provided to the traditional sector. For some ideas along these lines see, for example, Myint (2001), Galor et al. (2009), Acemoglu and Robinson (2012, pp. 258–271) and Mourmouras and Rangazas (2013b).

Sects. 7.3, 7.4, and 7.5 discusses the connection between missing markets and labor mobility. Sections 7.6 and 7.7 provide empirical applications of the model.

7.2 Perfectly Competitive Markets in the Traditional Sector

There are several explanations for the wage gap that do not rely on incomplete markets in the traditional sector. Each likely plays some role in explaining the gap, but each also has limitations. More empirical research is needed to determine which of the various explanations, including those that rely on incomplete markets, are the most generally applicable and the most important quantitatively. However, we believe the evidence suggests that the complete market approaches taken thus far cannot fully explain large wage gaps.

7.2.1 Urban Unemployment

One way to explain a relatively low wage in rural areas of developing countries is to assume that labor markets exist and clear in the traditional sector, but that non-competitive wage setting occurs in modern urban labor markets, say due to powerful unions, resulting in unemployment (Harris and Todaro 1970; Stiglitz 1974; Calvo 1978). The probability of being unemployed must be taken into account by those choosing to locate and work in the urban sector. A worker would only consider seeking employment in the urban sector if the wage there were high enough to compensate for the probability of being unemployment for some period. Thus, a wage-gap between the modern and traditional sectors is necessary for equilibrium where workers locate in both sectors.[3]

There are several difficulties with this theory. Unemployment in urban sectors of most poor countries is not particularly high—certainly not high enough to explain large wage gaps (Rosenzweig 1988; Caselli 2005). There is also disguised or unmeasured unemployment in rural sectors (Stiglitz 1988; Mazumdar 1989), suggesting *urban* unemployment is less of a problem when viewed *relative* to unemployment in rural areas.

One way around the absence of high unemployment rates in urban areas, or more precisely the absence of large unemployment rate *differentials* between urban and rural sectors, is to assume that there is an informal or traditional sector located in urban areas. In the informal-urban sector, workers are employed but at lower wages than they would receive in traditional agriculture. They are willing to endure this situation temporarily as they wait for higher-paying government or union jobs in the formal sector to materialize. In this way, non-competitive wage setting in the formal

[3]See Temple (2005) and Wang and Piesse (2011) for recent discussions of dual economies that focus on urban unemployment.

labor market of urban areas could still drive an equilibrium wedge between wages in urban and rural areas without the presence of relatively high unemployment rates in the urban sector.

However, there is evidence suggesting that the informal-urban sector is not simply a "waiting station" for formal-sector employment. In many cases the informal sector provides a lifetime employment choice that runs parallel to the possibility of working in the formal sector. As in the rural sector, informal urban workers receive lower wages than those in the formal urban sector. As compensation for the low wages, informal workers expect to eventually find greater rewards in the informal sector through some kind of entrepreneurial activity later in their life when they become owners or managers of informal businesses (Mazumdar 1989; Roseznweig 1988). This situation closely resembles a farm worker waiting to eventually take over the family farm, or otherwise own and operate his own small farm. The close parallel between the urban and rural circumstances of an informal worker provides further reason to focus on a formal-informal wage gap, rather than strictly an urban-rural wage gap.[4]

7.2.2 Human Capital Gaps

The formal sector, which by definition uses relatively advanced technologies and organizational structure, has relatively high payoffs to education (Herrendorf and Schoellman 2014; Vollrath 2014). Caselli and Coleman (2001) explain the declining wage gaps in the US over the twentieth century by assuming that education only enhances productivity "off the farm." They argue that at the beginning of the twentieth century high costs of education kept workers from leaving the farm for industry jobs. This caused the supply of qualified workers in industry to be low, creating a large education wage-premium, which in their model also doubles as a large nonfarm wage-premium. As the cost of education fell over the century, and the relative supply of workers in industry expanded, much of the wage gap was eliminated.

While the payoffs to education might be higher in industry, there is much evidence that they are also significant in agriculture—even in traditional agricultural settings (Schultz 1964; Harris 1972; Feder et al. 1982; Foster and Rosenzweig 1996; Goldin and Katz 2000). In addition, some of the empirical literature on the wage gap adjusts for experience and education differences and continues to find that a wage gap exists. Mazumdar (1989, Appendix A) finds that wages increase with firm size (an indirect measure of formal production), for *given* levels of education and experience. Jenkins and Knight (2002, Table 4.6) find large wage gaps across rural and urban sectors in Zimbabwe for workers with the *same* years of schooling. In

[4]The World Bank (1995) reports that 76% of workers in Sub-Saharan Africa are found in the informal sector—55% are agricultural workers that own their own farms or work in village farms and 21% are workers in services and industry.

Europe and the U.S. during the nineteenth century there were relatively small wage gaps *per hour* worked, suggesting that skill levels of workers in agriculture and non-agriculture were similar and that the payoffs to skill were not very different across sectors (Mourmouras and Rangazas 2009). However, as we shall see in Chap. 8, wages per worker may nevertheless be large because of differences in annual hours worked per worker across sectors.

Years of schooling for children of households in the modern sector are typically higher than those in the traditional sector (Cordoba and Ripoll 2006; Gollin et al. 2014; Vollrath 2009a, b, 2013). There is some suggestion that the "quality" of schooling is also different across the two sectors, as rural schools in developing countries are less well equipped and have fewer days of attendance over the course of a school year (e.g. Banerjee and Duflo 2011, Chap. 4; World Bank 2013). The quality differences may explain some of the differences in the rate of return to years of schooling.

We show in Sect. 7.4 that human capital gaps can also create annual wage gaps by influencing fertility and hours worked across sectors. Looking at the direct effect of education differences on productivity per hour worked alone may understate the role of human capital in explaining annual productivity gaps because of the possible connection between schooling and hours worked. On the other hand, the human capital gaps themselves may be due to missing markets and the dual economy structure of poor countries. We will discuss the connection between missing markets and human capital in more depth in Sect. 7.7.

7.2.3 Unmeasured Home Production

At least a portion of the wage gap between non-agriculture and agriculture is due to measurement error and non-traded goods. Workers in rural areas spend a larger portion of their work day producing unmeasured goods that are consumed at home. Mueller (1984) finds that agricultural workers in Botswana devote less than one hour per day to wage labor and the trading, vending, and processing of goods. If these workers are measured as a full unit of labor and their output is measured based only on market transactions, then their measured productivity is much less than their actual productivity.

The measured output of workers in agriculture is likely below the actual output, but it is unclear how large the measurement error is and how much it impacts the productivity gap. As mentioned above, survey data on consumption differences, which would include non-market production, nevertheless imply there are large wage gaps left to be explained.

7.2.4 Taxes, Fees, and Migration Costs

Workers migrating to the modern sector are likely to face many costs not found in the rural sector including taxes, discriminatory housing costs, union fees and the one-time cost of moving to the city. These migration costs imply that there must be a higher gross wage for workers to be willing to choose occupations in the urban sector.

Restuccia et al. (2008) assume that migration costs are the sole reason for the productivity gap between sectors. They focus on the productivity gap between agricultural and non-agricultural workers in currently developing countries. Similar to the estimates reported in Table 7.1, their estimates indicate that nonagricultural workers in poor countries are 6.7 more productive than workers in agriculture. To explain a 6.7-fold wage gap requires migration costs to be equivalent to an implicit tax on wages in the formal sector of 85% The necessary implicit “tax rate” on nonagricultural workers would have to be even higher if one accounts for the fact that (i) marketed agricultural goods are also taxed, (ii) the relevant tax rate should be *net* of any transfers or services generated from the tax revenue, and (iii) shared tenancy arrangements in agriculture carry implicit marginal tax rates. In China, for example, the government’s net tax rate on rural households is much *larger* than for urban households (Wang and Piesse 2010). There are certainly migration costs that limit the pace of migration out of traditional agriculture, but these largely one-time costs seem much too small to explain large and permanent gaps in wages.

7.2.5 Summary

From our perspective, the literature that assumes complete markets in the traditional sector cannot adequately explain the large wage and productivity gaps that persist in the early stages of development. In Sects. 7.3 and 7.4 we focus on explanations that assume the absence of markets in the traditional sector. A major advantage of this approach is that it offers a single unified explanation of gaps in wages, fertility, and saving rates across sectors.

7.3 Missing Land Markets in the Traditional Sector

This section extends the one-sector neoclassical growth model from Chap. 5 to a dual economy setting, with traditional and modern sectors. The key assumption creating the dual structure is that, in the absence of formal markets for land in the traditional sector, the children of traditional producers inherit the rights to operate a family farm from their parents. As is common in the literature, we assume farm land, or the right to farm the land under a tribal or tenancy agreement, is simply passed from one generation of farmers to the next (e.g. Bertocchi 2006; Doepke 2004; Drazen and Eckstein 1988; Galor et al. 2009; Hayashi and Prescott 2008).

We go one step further and assume the property rights over the farm are passed *only if* the children agree to operate it themselves—this ties the inheritance/management of traditional assets to occupational choice. Without an inheritance that is conditional on the recipient working and then managing the farm, family-based or traditional farming alone cannot explain why wage gaps persist in equilibrium.

To see this last point, suppose that the rights to family or village assets are not inherited. Instead, assume that all land is purchased in land markets by younger households and sold by older households (as in, for example, Chap. 6 and Hansen and Prescott 2002). In this case there is no compensation for low wages in the traditional sector and no reason for workers to accept a wage gap between sectors. In the absence of moving costs and mobility restrictions, migration of labor across sectors would completely eliminate any gap in wages. Alternatively, suppose land markets do not exist but the recipient of the land bequest could own the farm in absentia, without directly working and managing the farm. Here again there is nothing to prevent migration to the high-wage sector, so wage gaps cannot persist.

An approach based on a bequest that is conditional on the recipient working and managing the land is consistent with the strong tradition of family farming in the U.S. The link between the farm and the family in the U.S. was surprisingly strong well into the twentieth century. At the turn of the century, the farm in the U.S. was still largely operated individually and organized around the family and most of farm labor was provided within the family. Even by 1930 only 42% of all farms reported hiring labor outside the family (Ely and Wehrwein 1940, p. 162). As late as 1978, traditional family farms represented 88% of all farms, accounting for 63% of total farm production (Gardner 2002, pp. 56–57).

Intergenerational links to farming seem to be a common feature of developing countries more generally. Hayashi and Prescott (2008) claim that Japan's development was slowed by a social convention of passing the farm across generations within the family. Their paper includes a passage (pp. 605–606), suggested by Andrew Foster, arguing that the social convention may be strong enough to operate even in the presence of a land market. "First, the heir could sell the inherited farmland and live in the city to collect the higher urban income. However, to prevent this, the father could require the son to remain on the farm until he inherits the land. By the time his son inherits the estate, it may be too late for him to start a career in the city." Consistent with Foster's explanation, Collier et al. (1986) find that those individuals who indefinitely migrate away from farms in Tanzania tend to lose their land entitlement. In China, explicit migration rules cause migrants to the city to give up ownership claims to land and small businesses in rural areas (Au and Henderson 2006). Basu (1991, pp. 24, 131–133) discusses the difficulties of establishing property rights over land and the lack of land sales in rural India.

One could also broaden the notion of family inheritance, beyond land, to include inheriting the *skills* to operate a family farm or more generally a small family business (Lord and Rangazas 2006). For example, Hong Kong and Singapore went through the same coincident economic transformations and demographic transitions as other developing countries, without having economies that were based on agriculture to any significant degree. Despite the absence of an agricultural sector, there was nevertheless an economic transformation from informal home-based production to formal firm-based production. In Hong Kong, as late as 1971, 69% of all manufacturing establishments were located in domestic premises. By 1978, as part of their "Growth Miracle," that share fell to 44% (Young 1992, p. 19).[5]

7.3.1 Modern Sector

The modeling of the modern sector is similar to the one-sector model in Chap. 5. For the reader's convenience, we provide a quick summary of its main features here.

Producers in the modern sector are neoclassical firms with the standard Cobb-Douglas production function

$$Y_t = AK_t^{\alpha} H_t^{1-\alpha}, \tag{7.1}$$

where Y denotes output, K is the physical capital stock, H is the effective labor supply in the modern sector, A is sector-specific TFP, and α is the capital share parameter. The effective labor supply is endogenously determined by population size, the fraction of the working population in the modern sector, and the schooling received by workers.

The firms operate in perfectly competitive markets for goods and inputs. The standard profit-maximizing factor-price equations for the rental rates on effective labor supply and physical capital are

$$w_t = A(1-\alpha)k_t^{\alpha} \tag{7.2a}$$

$$r_t = \alpha A k_t^{\alpha-1}, \tag{7.2b}$$

where r is the rental rate for capital and $k = K/H$. Note, for simplicity, we assume $D_t \equiv 1$ in this chapter.

Fundamental household characteristics do not differ by sector. All households live for two periods; one period of childhood and one period of adulthood. Households value family consumption (c_t) and the adult human capital (h_{t+1}) of all their children (n_{t+1}). Preferences are given by

$$U_t = \ln c_t + \psi \ln (n_{t+1} h_{t+1}),$$

where $\psi > 0$ is a preference parameter.[6] This preference specification is a simple way of capturing the idea that parents value both the quantity and the quality of children.

---

[5]Focusing on this interpretation—the inheritance of specific human capital gained from experience and word of mouth—may be a way of explaining wage gaps in the presence of land markets. We explore this possibility in later chapters.

[6]In Chap. 8 we show how one can explicitly model the preference to pass traditional assets to the next generation, with and without land markets. Being explicit about this preference adds notation without altering any qualitative conclusions, so we treat it here as a social convention, that is not explicitly linked to preferences, as has typically been done in the literature (e.g. Drazen and Eckstein 1988; Bertocchi 2006; Hayashi and Prescott 2008; Galor et al. 2009).

It has been used extensively in the literature on fertility and growth (e.g., Galor and Weil 2000; Greenwood and Seshadri 2002; Hazan and Berdugo 2002; Moav 2005). For simplicity only, in this chapter we avoid using a life-cycle assumption with two periods of adulthood or a bequest motive to generate domestic household saving. Instead we assume a small open economy with perfect capital mobility. All physical capital is owned by foreign households.[7]

Adults in elastically supply one unit of labor. In Sect. 7.4 we relax this assumption and consider differences in work time as a source of the productivity and wage gaps across sectors. Children have an endowment of $T < 1$ units of time that they can use to attend school (e_t) or work ($T - e_t$).[8] Children have less than one unit of time to spend productively because in the very beginning years of childhood they are too young to work and in older childhood they do not have the mental or physical endurance to work as long as an adult.

While children may work as they become older, they are also expensive to care for and feed. To raise each child requires a loss of adult consumption equal to a fixed fraction η of the adult's first period wages. It simplifies subsequent expressions we encounter to explicitly assume the loss in adult consumption is due to a loss in adult work time needed to raise children. Each child causes parents to sacrifice η of the full unit available for work.

Very young children are not productive but receive some early education. So each child invests at least $\bar{e}$ units of time into learning during the first portion of their childhood. This gives older children $\gamma\bar{h}_t = \gamma\bar{e}^{\theta}$ units of human capital that can be used in production during the later years of childhood, where $0 < \theta < 1$ is a parameter that gauges the effect of schooling on human capital accumulation and $0 < \gamma < 1$ reflects the fact that children lack relative physical strength or experience in applying knowledge to production compared to an adult. Adult human capital of the same person in the next period is $h_{t+1} = e_t^{\theta}$. Thus, a person is more productive in adulthood than in childhood because of greater strength and experience ($1 > \gamma$) and additional schooling ($e_t \geq \bar{e}$).

The household maximizes utility subject to the budget constraint,

$$c_t + n_{t+1}\eta w_t h_t = w_t h_t + n_{t+1} w_t \gamma \bar{h}(T - e_t).$$

In addition to the standard first order condition for consumption, the choices of n_{t+1} and e_t yield the following first order conditions

[7]See Chaps. 8 and 9 for the more complicated closed economy case where domestic saving finances physical capital accumulation in a dual economy.

[8]The variable e captures all the time spent in school during childhood and thus is a function of the "quantity" of schooling, years, and also one dimension of the "quality" of schooling, the hours attended within a year, that has been stressed by Banerjee and Duflo (2011), Cordoba and Ripoll (2006), and Rangazas (2000, 2002).

$$\frac{\psi\theta}{e_t} \leq \lambda_t n_{t+1} w_t \gamma \bar{h} \tag{7.3a}$$

$$\frac{\psi}{n_{t+1}} = \lambda_t [\eta w_t h_t - (T - e_t) w_t \gamma \bar{h}], \tag{7.3b}$$

where λ_t is the Lagrange multiplier.

Equation (7.3a) says the marginal utility of additional child quality must be equated to the marginal value of consumption lost from allowing children of working age to attend school. The strict inequality holds when the marginal cost of educating children beyond the schooling received in their early years, $\bar{e}$, exceeds the marginal benefit. In this case, parents are content to set $e_t = \bar{e}$.

Equation (7.3b) says the marginal utility of additional children must be equated to the marginal value of lost consumption. Consumption is lost from having additional children because we assume the cost of children exceeds the earnings that older children bring to the household.

Solving the model gives us the following demand functions for children and schooling

$$n_{t+1} = \frac{\psi}{(1+\psi)\left(\eta - \gamma(T - e_t)\left(\bar{e}/e_{t-1}\right)^{\theta}\right)} \tag{7.4a}$$

$$e_t = \max\left[\frac{\theta\left(\eta\left(e_{t-1}/\bar{e}\right)^{\theta} - \gamma T\right)}{\gamma(1-\theta)}, \bar{e}\right]. \tag{7.4b}$$

Assuming that e_{t-1} is sufficiently high, an assumption that we make throughout, a dynamic interaction results that contributes to economic growth and the demographic transition.[9] Greater schooling raises adult earnings relative to older children's earnings. This raises the net cost of having children, so fertility declines. Reduced fertility and greater consumption lowers both the size and the value of forgone earnings from schooling children, encouraging a further rise in schooling. Thus, the sole factor driving fertility down is the rise in schooling.

7.3.2 Traditional Sector

The traditional sector differs from the modern sector because of differences in the production technology and because the technology and land used in production is inherited and operated by a family—creating lifetime entrepreneurial or residual

[9]In Chaps. 4 and 5, we focus on the possibility of a poverty trap associated with initial conditions. If parents' education is not sufficiently high, the incentive to educate children beyond their early years is missing. Different initial conditions create schooling differences across sectors that are one factor generating persistent productivity gaps. Here we focus primarily on other sources of the gap.

income to supplement wages. For simplicity we assume that the same goods are produced in each sector. This is a common assumption in the dual economy approach when the specific focus is not on the relative price of the different goods produced in the two sectors. While the framework is general enough to interpret traditional production as representing any informal business, we refer to traditional firms as "family farms."

Traditional output is produced using the following Cobb-Douglas technology

$$O_t = \tilde{A} l_t^{\rho} f_t^{1-\rho}, \tag{7.5}$$

where O output, l is land per farm, f is effective farm labor per household in the traditional sector, $\tilde{A}$ is sector-specific TFP, and $0 < \rho < 1$ is a technology parameter. The technology differs from that used in industry because land is an input rather than physical capital. This assumption is meant to capture the idea that production in the traditional sector does not rely heavily on the plant and equipment used in the "factories" of the modern sector. In general, the labor-share parameter may also differ across sectors.[10]

There is no market for the land, either to buy or rent. For now we will assume that there is a market for farm labor, so that households could hire labor beyond that provided by the family (although in equilibrium, because all households are identical, this will not happen). In Sect. 7.4, we consider how things change if there is no labor market in the traditional sector. We assume that the young household inherits the skills and land needed to operate a farm from their parents.

The presence of residual income from the inherited family farm and the fact that the rental rate on effective labor earned working on family farms ($\tilde{w}$) will in general differ from that earned working for firms (w) are two important differences between farming and working in the modern sector. Although not our main focus, we can also consider a third difference. Schooling in the traditional sector may lag behind schooling in the industrial sector. This could be because schooling became available in urban areas of the modern sector before the rural areas of the traditional sector or because it was generally harder to enforce child labor/mandatory schooling laws in the traditional sector. If initial schooling is lower in the traditional sector then schooling in the traditional sector will lag behind, via the dynamics of (7.4b), for every subsequent generation. These three potential differences will in general cause all farm choice variables to differ from those chosen in the modern sector.

The lifetime budget constraint for a household choosing to follow their parents and work in the traditional sector is

$$\tilde{c}_t + \tilde{n}_{t+1} \eta \tilde{w}_t \tilde{h}_t = \tilde{w}_t \tilde{h}_t + \tilde{n}_{t+1} \tilde{w}_t \gamma \bar{h} \left(T - \tilde{e}_t\right) + \left(O_t - \tilde{w}_t f_t\right), \tag{7.6}$$

[10]We could also introduce labor-augmenting technology and allow each to differ across sectors, but this would not affect our major points.

where $O_t - \tilde{w}_t f_t$ is the residual income from farming. Maximizing utility subject to (7.5) and (7.6) yields the following demands for children, schooling, and farm labor

$$\tilde{n}_{t+1} = \left(\frac{\psi}{1+\psi}\right) \frac{1 + (\rho/(1-\rho))\left(f_t/\tilde{h}_t\right)}{\eta - \gamma\left(T - \tilde{e}_t\right)\left(\bar{e}/\tilde{e}_{t-1}\right)^{\theta}} \tag{7.7a}$$

$$\tilde{e}_t = \max\left[\frac{\theta\left(\eta\left(\tilde{e}_{t-1}/\bar{e}\right)^{\theta} - \gamma T\right)}{\gamma(1-\theta)}, \bar{e}\right] \tag{7.7b}$$

$$f_t = \left[\frac{1-\rho}{\tilde{w}_t}\right]^{\frac{1}{\rho}} l_t \tag{7.7c}$$

Contrasting (7.7a) to (7.4a), reveals that production on the family farm introduces a new term in the fertility demand function, $\rho f_t/(1-\rho)\tilde{h}_t$, that raises fertility (other things constant). The numerator of this term is residual farm income, which can be shown to be proportional to the labor employed at the farm. The denominator is the potential "full" wage that can be earned as a worker, which determines the opportunity cost of having children. The more important family production is, relative to the opportunity cost of children, the stronger is the demand for children. This is not a pure wealth effect, but rather is an effect that arises when one form of wealth, residual income from the ownership of family production that does *not* affect the net cost of children, increases relative to another form of wealth, adult earnings from work effort, that *does* affect the net cost of children. A shift in the composition of family wealth away from family production and toward adult wages causes the net cost of children to rise, for a given level of total family wealth, and the demand of children falls.

The demand for schooling takes the same form as in (7.4b), although there may be different initial conditions for households living in the traditional sector. There is no effect of family production on schooling because of two offsetting effects. To see these effects, first note that fertility raises the cost of schooling children (more children means greater forgone consumption of parents as schooling rises and child labor income falls). Second, note that the level of parental consumption determines the marginal *value* of forgone consumption associated with greater schooling (higher parental consumption levels means parents can better "afford" the lost consumption associated with more schooling). Family production raises *both* fertility and parental consumption, other things constant. As just mentioned, higher fertility lowers the incentive to school children, but a higher consumption level raises the incentive to school children. With our functional forms for preferences and human capital production, these two effects always exactly offset.

The demand for labor in (7.7c) results from the farm owner hiring labor to equate the marginal product of effective labor to the agricultural rental rate on human capital, perfectly analogous to the demand for labor by neoclassical firms in competitive factor markets. Note that we allow the demand for labor at an individual farm to be is less or greater than the supply of labor coming from the household owning the farm. From the perspective of individual farms, some of the household labor may have to be supplied to neighboring farms or, alternatively, farm "hands" may have to be hired to supplement family labor supply.

7.3.3 Equilibrium

In a small open economy, the return to physical capital must equal the exogenous international rental rate, r. Equating r_{t+1} to r in (7.2b), allows one to solve for the equilibrium value k. Eq. (7.2a) can then be used to solve for the equilibrium w_t.

A household born into the traditional sector has the option of staying there (receiving land and then farming it) or of working for a firm in the modern sector. Those who were born in families that work in the modern sector must stay there because they have no possibility of obtaining land. As discussed, fertility is higher in the traditional sector, so there must be some movement of the population from traditional to the modern sector, otherwise the fraction of workers in the modern sector would *fall* over time (contradicting the economic transformation associated with development). For there to be some movement, but not a *complete* shift of the population to the modern sector, the human capital rental-rate gap between the two sectors must be such that a worker born into the traditional sector is indifferent about staying there.

Calculating the indirect utility function of the household under each option and then equating them, gives an expression for the equilibrium human capital rental-rate gap that makes traditional households indifferent about their locational/occupational choice

$$\frac{w_t}{\tilde{w}_t} = \left[1 + \frac{\rho f_t}{(1-\rho)\tilde{h}_t}\right]^{1+\psi}. \tag{7.8}$$

The term in the square-bracket on the right-hand-side exceeds one due to the residual income received from operating a family farm. Rental rates in modern sector must exceed those in the traditional sector to compensate for giving up the farm and its additional source of income.

The equilibrium rental-rate gap is further widened by the fact that the exponent on the term in square bracket exceeds one. With endogenous fertility, being indifferent about working in the two sectors requires more than the equality of lifetime resources. The higher wages in the modern sector increase the opportunity cost of time spent away from work. Thus, working in the modern sector raises the cost of having and raising children relative to working in the traditional sector. This results in fewer children for nonfarm households. Given that parents like children ($\psi > 0$), they must receive a wage premium in the modern sector that compensates them for the fewer children they will have.

Finally, we must ensure that the labor market in the traditional sector clears. Let $\tilde{N}_t$ denote the number of households choosing the traditional sector. The aggregate demand for farm labor is $\tilde{N}_t f_t$. Each household supplies their own labor and that of their children to the market for farm labor. The aggregate supply of effective labor to the traditional sector is $\tilde{N}_t \tilde{h}_t \left[1 - \eta \tilde{n}_{t+1} + \tilde{n}_{t+1} \gamma \frac{\bar{h}}{\tilde{h}_t}(T - \tilde{e}_t)\right]$. All households that remain in the traditional sector are identical, thus in equilibrium each household must demand enough labor to absorb the quantity of labor they supply,

$$f_t = \tilde{h}_t \left[1 - \eta \tilde{n}_{t+1} + \tilde{n}_{t+1} \gamma \frac{\bar{h}}{\tilde{h}_t}(T - \tilde{e}_t)\right]. \tag{7.9}$$

Using (7.7a), we can write

$$f_t = \tilde{h}_t \left[1 - \tilde{n}_{t+1}\left(\eta - \gamma \frac{\bar{h}}{\tilde{h}_t}(T - \tilde{e}_t)\right)\right] = \tilde{h}_t \left[1 - \frac{\psi}{1+\psi}\left(1 + \frac{\rho}{1-\rho}\frac{f_t}{\tilde{h}_t}\right)\right].$$

Solving for $f_t/\tilde{h}_t$ gives us.

$$\frac{f_t}{\tilde{h}_t} = \frac{1-\rho}{1+\psi-\rho}.$$

The aggregate quantity of raw land is fixed at L. The quantity of land per farmer is then $l_t = L/\tilde{N}_t$. Thus, l_t falls over time as long as the population of farming households increases over time.

Equations (7.7a, 7.7b, 7.7c 7.8, and 7.9) determine the equilibrium paths of $\tilde{e}_t, \tilde{n}_{t+1}, \tilde{w}_t, f_t$, and $\tilde{N}_t$. In particular, using (7.7c) the number of farmers is given by

$$\tilde{N}_t = L \frac{[(1-\rho)]^{\frac{1}{\rho}}}{f_t \tilde{w}_t^{1/\rho}}.$$

The number of farmers is decreasing in both $\tilde{w}$ (the cost of farm labor) and the effective supply of farm labor from a farm family (to absorb a larger effective supply of family workers, more land per farmer is needed to generate a rise in labor productivity and a greater demand for labor).

The supply of *potential* farmers in period-t includes the entire population of children of the previous generation of farmers, $\tilde{n}_t \tilde{N}_{t-1}$. In equilibrium we must have $\tilde{N}_t \leq \tilde{n}_t \tilde{N}_{t-1}$. Along a transitional growth path, we will have $\tilde{N}_t < \tilde{n}_t \tilde{N}_{t-1}$, so that a fraction of the potential farmers leave for the modern sector each period. This implies that there are actually three types of households: (i) the original dynastic line of modern sector households, (ii) modern sector households that have migrated from the traditional sector, and (iii) households remaining in the traditional sector.

In general, the initial schooling of types (i) and (ii) will differ, causing the schooling of *all* households in each of their dynastic lines to differ during the transition. The schooling difference causes fertility of the migrant households to be different from a household whose dynastic line originated in the modern sector area. The migrant household's fertility is given by

$$\widehat{n}_{t+1} = \frac{\psi}{(1+\psi)\left(\eta - \gamma\left(T - \tilde{e}_t\right)\left(\bar{e}/\tilde{e}_{t-1}\right)^{\theta}\right)}. \tag{7.10}$$

Although, the migrant's schooling is the same as those that remain in the traditional sector, their fertility is lower because of the absence of residual income from operating the family farm. Thus, the modern sector will contain two distinct types—both of which differ from the household remaining in the traditional sector, along at least one dimension.

Both sources of the wage gap, rents from inherited land and schooling differences, help explain higher fertility in the traditional sector. Average fertility in the economy falls as schooling rises in both sectors and as households migrate from the traditional to the modern sector. The fraction of the population in the traditional sector shrinks as the effective labor supply and the cost of labor rises in the traditional sector.

Note that because fertility is higher in the traditional sector than it is for migrants, consumption for traditional households must be lower than for migrants in order for utility of the two households to be equal. If urban natives have greater schooling and human capital than the migrants, then they will have higher potential income. Since consumption is a normal good, this implies that they will also have higher consumption than migrants. Thus consumption will be highest for urban natives, lowest for traditional households, with consumption of migrant households in the middle.

During the structural transformation and demographic transition of the economy, the wage gap due to differences in human capital rental rates remains constant. To see this, substitute the solution for $f_t/\tilde{h}_t$ from (7.9) into (7.8) to write the human capital rental rate gap as

$$\frac{w_t}{\tilde{w}_t} = \left[\frac{1+\psi}{1+\psi-\rho}\right]^{1+\psi}. \tag{7.11}$$

7.3.4 Summary

With an absence of land markets in the traditional sector, there is a greater tendency for land and family farms to be passed from one generation to the next. Historical and cross-country observations suggest that an important condition for receiving an inheritance of this type is that children remain in the traditional sector to gain the experience needed to operate the farm and "keep it in the family." This institutional arrangement ties occupational choice and geographic location to land ownership in the traditional sector. In equilibrium, traditional workers are willing to take lower wages than those offered in the modern sector because they are compensated by rental income on the land they eventually own and by their larger families.

7.4 Missing Labor Markets in the Traditional Sector

Chapter 8 documents that in the historical development of Europe and the U.S., most of the measured annual wage gap between agriculture and non-agriculture was due to lower annual hours worked in agriculture rather than lower productivity per hour worked. In many developing countries today, differences in hours worked explain a significant portion of the annual gaps in worker productivity (Gollin et al. 2014; Vollrath 2013). In Chap. 8 we show how one can adjust the analysis of this Chapter to reinterpret the wage gap as a gap in hours worked. Our approach is to think of hours worked in the traditional sector as being *constrained* by seasonality, the physical demands of traditional agriculture, and the time spent securing property rights. Similar to 7.3, the loss in hours worked in the traditional sector is compensated in equilibrium by rents from inherited land and from a larger family.

Vollrath (2009b) takes a different approach and explains the shortfall in agricultural hours as a *choice* made by traditional farmers to work less than workers in the modern sector. This section follows Vollrath's lead and allows for an endogenous choice of work effort. An important assumption, needed to generate lower work effort as a *choice* of farmers, is the absence of labor markets in the traditional sector. We begin with a simple case where work effort is the only choice variable and then extend the analysis to include fertility and schooling.

7.4.1 Endogenous Work

In the modern sector, we continue to assume perfectly competitive markets exist for labor and capital as in Sect. 7.3. We also continue to assume that the international market for capital determines the return to capital, r, that the small open economy takes as given. The competitive wage rate is $w_t = (1-\alpha)Ak^{\alpha}$, where k is the physical capital intensity that satisfies (7.2b) when $r_t = r$.

As before, each household is endowed with one unit of time but now they can choose either to work, z_t, or to consume leisure (or engage in home production), $1 - z_t$. The modern sector household chooses z_t to maximize $\ln c_t + \zeta \ln(\nu + 1 - z_t)$, where ζ and ν are preference parameters, subject to the budget constraint $c_t = w_t z_t$. The solution is $z_t = \frac{1+\nu}{1+\zeta}$. We set $\zeta = \nu$, so that the modern household chooses to work the full unit of time, as was assumed in Sect. 7.3.

Households in the traditional sector have preferences and skills identical to those in the modern sector. As in Sect. 7.3, assume that there is no market for the buying and selling of land and each traditional household inherits l_t from their parents (or is granted the property rights to farm a portion of village land holdings).[11] In addition, unlike Sect. 7.3, we now assume that there is no labor market in the traditional sector. The farm household is the only supplier of labor to the farm. The budget constraint of the traditional household is then $\tilde{c}_t = l_t^{\rho}\left(\tilde{z}_t\right)^{1-\rho}$, where $\tilde{z}$ is the labor supply choice of the individual farmer.

The solution to the traditional household's problem is $\tilde{z}_t = (1-\rho)/\left(1-\frac{\rho}{1+\upsilon}\right) < 1$, where $\zeta \equiv \nu > 0$. As in Vollrath (2009b), the traditional household chooses less work effort than the modern sector household because the diminishing marginal product of labor on the family farm lowers the marginal benefit of work relative to the situation where the marginal reward to work is determined by a perfectly competitive wage rate that is independent of the household's work choice.

The fact that the traditional household works less and enjoys more leisure, implies that in equilibrium, in order to be indifferent about locational choice, the traditional household must generate less income and consumption. Thus, modern sector households have greater income, (annual) labor productivity, and consumption, all of which are compensation for taking less leisure.

7.4.2 Schooling and Fertility

We now show that the argument above extends to the situation where the quantity and quality of children are also household choice variables. The extended choice problem creates an interaction between schooling, fertility, and work effort that increases the wage gaps across sectors.

Now let total time be allocated across work, leisure/home production, and rearing children, so that leisure time is $1 - z_t - \eta n_{t+1}$. The modern sector household chooses z_t, e_t, and n_{t+1} to maximize $\ln c_t + \zeta \ln(\nu + 1 - z_t - \eta n_{t+1}) + \psi \ln h_{t+1} n_{t+1}$, subject to

$c_t = w_t h_t z_t + n_{t+1} w_t \gamma \bar{h}(T - e_t)$. The solution is

$$e_t = \max\left[\frac{\theta\left(\eta\left(e_{t-1}/\bar{e}\right)^{\theta} - \gamma T\right)}{\gamma(1-\theta)}, \bar{e}\right] \tag{7.12a}$$

[11]The key assumption needed to generate the shortfall of work effort in the traditional sector is the absence of a labor market. Vollrath (2009b) shows the result holds in the presence of a *rental* market for land provided the farmers themselves own the land and rent it to each other rather than from absentee landlords. The situation where farmers are the predominant owners of land is more likely where there is no *asset* market for land and the ownership of land is passed down from one generation of farmers to the next.

$$n_{t+1} = \frac{\psi(1-\theta)(1+\nu)}{\eta(1+\psi+\zeta)} \frac{1}{\left(1-(\gamma T/\eta)\left(\bar{e}/e_{t-1}\right)^{\theta}\right)} \tag{7.12b}$$

$$z_t = \frac{(1+\nu)}{1+\frac{\psi(1-\theta)}{1+\theta\psi-(1+\psi)(\gamma T/\eta)\left(\bar{e}/e_{t-1}\right)^{\theta}}\left(1+\frac{\zeta}{\psi(1-\theta)}\left(1-(\gamma T/\eta)\left(\bar{e}/e_{t-1}\right)^{\theta}\right)\right)}. \tag{7.12c}$$

The schooling equation is identical to that derived in Sect. 7.3. Although the form of the fertility equation is somewhat different, fertility is declining in schooling as before. The main new wrinkle is that now there is a connection between education, fertility, and work. One can show that as e_{t-1} increases, and adult human capital rises, the optimal level of work effort in (7.12c) also rises.

In most settings with log preferences, the rise in the reward to work has offsetting income and substitution effects that leave work effort unchanged. Here, as human capital rises and fertility falls, time available for leisure increases. The freed leisure time causes a decline in the marginal value of time and lowers the cost of market work. Thus, fewer children allows more time for both leisure/home production and market work.

The traditional household has the same preferences as the modern sector household and chooses $\tilde{z}_t$, $\tilde{e}_t$, and $\tilde{n}_{t+1}$ to maximize $\ln \tilde{c}_t + \zeta \ln\left(\nu + 1 - \tilde{z}_t - \eta \tilde{n}_{t+1}\right) + \psi \ln \tilde{h}_{t+1}\tilde{n}_{t+1}$, subject to $\tilde{c}_t = l_t^{\rho}\left(\left(\tilde{h}_t\tilde{z}_t + \tilde{n}_{t+1}\gamma\bar{h}\left(T-\tilde{e}_t\right)\right)\right)^{1-\rho}$, where family farm labor now includes the work time of the household's older children. The solution to the traditional household's problem is

$$\tilde{e}_t = \max\left[\frac{\theta\left(\eta\left(\tilde{e}_{t-1}/\bar{e}\right)^{\theta}-\gamma T\right)}{\gamma(1-\theta)}, \bar{e}\right] \tag{7.13a}$$

$$\tilde{n}_{t+1} = \frac{\widetilde{\psi}(1-\theta)(1+\nu)}{\eta\left(1+\widetilde{\psi}+\widetilde{\zeta}\right)} \frac{1}{\left(1-(\gamma T/\eta)\left(\bar{e}/e_{t-1}\right)^{\theta}\right)} \tag{7.13b}$$

$$\tilde{z}_t = \frac{(1+\nu)}{1+\frac{\widetilde{\psi}(1-\theta)}{1+\theta\widetilde{\psi}-\left(1+\widetilde{\psi}\right)(\gamma T/\eta)\left(\bar{e}/\tilde{e}_{t-1}\right)^{\theta}}\left(1+\frac{\widetilde{\zeta}}{\widetilde{\psi}(1-\theta)}\left(1-(\gamma T/\eta)\left(\bar{e}/\tilde{e}_{t-1}\right)^{\theta}\right)\right)}, \tag{7.13c}$$

where $\widetilde{\psi} \equiv \psi/(1-\rho)$ and $\widetilde{\zeta} \equiv \zeta/(1-\rho)$.

The form of (7.13a, 7.13b, 7.13c) is the same as (7.12a, 7.12b, 7.12c), except that $\widetilde{\psi} > \psi$ and $\widetilde{\zeta} > \zeta$. The presence of diminishing marginal productivity to family labor $(1-\rho < 1)$ has an effect on behavior that is equivalent to raising the relative value of fertility and leisure time. One can show that this implies $\tilde{z}_t < z_t$ and $\tilde{n}_{t+1} > n_{t+1}$, so traditional households work less and have more children than modern sector households. This is true even if there is no difference in schooling across sectors. In equilibrium, this implies consumption and income of traditional households is lower than in modern households.

As discussed in the case of the modern sector household we know when e_{t-1} increases, and adult human capital rises, the optimal level of work effort also rises. The same dynamic is true of traditional households. Thus, any differences in schooling across sectors not only increases the productivity gap directly but also does so indirectly by creating a larger gap in market work effort. Just as schooling is one of the fundamental causes of fertility differences across sectors, schooling is also one of the causal factors in explaining differences in market hours.

7.4.3 Summary

The absence of a labor market in the traditional sector causes family producers to rely on family labor. The diminishing returns to work on the family farm lowers the reward to work relative to a market setting where the competitive market wage is independent of an individual's labor supply choice. The absence of a labor market therefore leads to less work in the traditional sector creating an annual wage gap due to differences in hours worked.

A difference in schooling across sectors widens the difference in hours worked. More schooling lowers fertility and frees time for leisure and work. Thus, greater schooling and lower fertility in the modern sector leads to greater hours of market work per adult worker and a larger annual wage gap. This particular effect applies even when there are complete markets in both sectors.

7.5 The Forces That Bind Us: Missing Markets and Labor Mobility

In this section we present additional discussion of the ways missing markets restrict labor mobility, in both historical and current development.

7.5.1 Missing Land Markets in Historical Development

Alan Macfarlane wrote an influential work on the early history of the English economy called the *Origins of English Individualism* (1978a). The study sought to determine the importance of the traditional sector, or what Macfarlane called the peasant economy, in English history. In a summary of his work, Macfarlane (1978b, p. 256) characterized the peasant economy as being centered around the family, in the same way we are thinking about the traditional sector in our dual economy model.

> For our purposes, the central feature is that ownership is not individualized. It was not the single individual who exclusively owned the productive resources, but rather the household. Thus, the heirs have as much right to the resources as the present "owners." The present occupants of the land are managers of an estate; they cannot disinherit their heirs, the father is merely the leader of a production team. This is an ideal-type situation, but something like it has been documented for many parts of the world up to this century, for instance in Eastern Europe, India, and China. In this situation, farm labor is family labor. Hired labor is almost totally absent. Production is mainly for use, rather than for exchange in a market. Cash is only occasionally used within the local community. Land is not viewed as a commodity which can be easily bought and sold. There is a strong emotional identification with a particular geographic area. Consequently, there is rather little geographic mobility: any movement to the towns is one-way, with few people returning to the countryside.

Macfarlane argued that while the traditional economy was important in most of Europe through the nineteenth century, it had almost disappeared in England centuries earlier. Even before the sixteenth century in England, he argues (1978b, p. 257).

> These again gave no hint that on any of the criteria elaborated above, England faintly resembled a peasant society. Land is a marketable commodity, geographic and social mobility is widespread, property is held by the individual rather than the family.

He cites visitors to England from Europe that noted this feature of the economy with surprise and concluded that "individualism" was unique to England. He also suggested that this helps explain why the Industrial Revolution started there, rather than in other countries of Europe.

Research motivated by Macfarlane's work has generally confirmed that land markets existed in England early on. There was clearly the legal right to disinherit and to buy and sell land freely. The question is the extent to which the legal rights were actually used. In a review of the recent evidence on the matter, French and Hoyle (2003, p.261) conclude that while land markets and legal rights to buy and sell existed, they were not used frequently.

> No one would deny that the English had the right to buy and sell land without reference to their kin (although they might give them the first option). However, we now have some indication of how often they sold their land and the answer is not often.

Thus, in traditional settings, even when land markets exist, most households behave as if they do not. This is evidently due to some preference to keep the land within the family. Assuming that land markets do not exist can be viewed as a modeling shortcut that includes situations where there literally are no markets, which certainly is the case in many settings, and where there are markets that are rarely utilized.

In Chap. 8, we show how one can model the preference to keep land or the farm in the family. In Chap. 9, we argue that it is not just the land or assets that bind the family together in traditional settings. In addition to possible physical inheritance, children gain specific human capital, local business knowledge, and community relationships when they work in the family business of the traditional sector. In short, they inherit the traditional sector technology.

7.5.2 Missing Land Markets in Currently Developing Countries

Janvry et al. (2015) study the effects of the absence of land titles in Mexico during the twentieth century. Land reforms over the century granted community members small plots of land. The plots lacked clear titles and ownership claims were closely linked to farmer usage. Owners that left the land idle or allowed others to farm the land risked the loss of property rights. The informal system of property rights tied owners/workers to small and often low productivity land.

In the late 20th and early 21st centuries, Mexico carried out a large scale certification program. The certification created tradable land titles, specifying the name of the owner for each plot of land. By the end of the program over 36 million farmers had received titles.

As predicted by our model, once work and land ownership was broken, Janvry et al. (2015) found increased migration out of rural areas. Their estimates suggest that land titles raised migration flows by about 20%. The migration outflows were concentrated in areas with low land productivity.

7.5.3 Missing Land Markets in Cities of Currently Developing Countries

As we mentioned in the introduction to this chapter, the traditional sector extends beyond rural areas. Many cities of developing countries have large informal sectors that share many of the characteristics of the rural traditional sector. One of the most important common characteristics is the absence of formal land markets. Field (2007) examines the effect of missing land markets on the labor supply of urban households.

The lack of formal property rights on land holdings forces urban households in the informal sector to expend resources guarding their property. Protecting their land holdings diverts labor time from the market and production. Filed estimates the magnitude of the effect on labor supply using data from a nationwide program in Peru that issued formal property titles to more than 1.2 million urban households between 1995 and 2003—the first major titling effort in the developing world. Her objective is to assess whether security improvements associated with titling increased hours worked in the formal labor market.

Field finds that households with no legal claim to property spend 13 h a week maintaining security. They are also 40% more likely to work at home rather than in the market. Overall, household labor supply was 16 h a week greater for households with titled property. Another consequence of titling was a reduction in child labor. Evidently, child labor is used to compensate for the lost market work of adults who spend time securing their untitled property.

7.5.4 Missing Credit and Insurance Markets in Current Development

Munshi and Rosenzweig (2009, 2016) provide a different example of how missing markets can lower labor mobility. They begin by acknowledging that the mechanisms we have been stressing generally serve to restrict labor mobility in developing countries (p. 1).

> Increased mobility is the hallmark of a developing economy. Although individuals might be tied to the land they were born on and the occupations that they inherit from their parents in a traditional economy, the emergence of markets allows individuals to seek jobs and locations that are best suited to their talents and abilities.

However, their main goal is to explain the unusually weak labor mobility in India. Their explanation is based on missing credit and insurance markets, rather than missing land markets.

India's labor mobility significantly lags that of other developing countries of similar size and state of development. In 1975, China, Indonesia, India, and Nigeria all had adult populations that were between 17 and 24% urbanized. In 2000, the urbanization rate was between 35 and 45% for China, Indonesia, and Nigeria, but was only about 27% for India. Generally, India's urbanization rate is about 15 percentage points lower than that of other developing countries with comparable per capita income. The permanent migration rate of Indian men out of their original village is less than 10%. The lack of mobility coincides with a persistent real urban-rural wage gap of more than 25%.

Munshi and Rosenzweig provide evidence that workers in India have low mobility because they are bound to their villages by the services provided by rural *Jati*-based communities. In the absence of formal credit and insurance markets, *Jatis* (tribes or clans) have been active in smoothing household consumption for centuries. The *Jati*- social network is effective in providing loans and insurance because of its ability to monitor, and potentially punish, behavior of local villagers. While households could instead obtain informal credit from moneylenders and employers, the cost tends to be much higher.

Thus, similar to missing land markets, missing credit and insurance markets can restrict the movement of labor out of the traditional sector. Relatively low-cost financial services in the village cause households to ignore better work opportunities in the modern sector.

7.6 Asian Growth Miracles

We can talk more explicitly about the connection between the structural transformation and growth by writing out the model's expression for worker productivity. Worker productivity is the total output of modern sector firms and traditional sector producers, divided by the total population of workers. The population of workers includes the older children that do not attend school. For simplicity, in writing out the average productivity expression, we assume equality between the labor shares ($\alpha = \rho$) and schooling ($e_t = \tilde{e}_t$) across sectors. Worker productivity is then

$$\frac{Y_t \;\; + \;\; \tilde{N}_t O_t}{N_t^* + N_{t+1}^*(T - e_t) + \tilde{N}_t + \tilde{N}_{t+1}(T - e_t)},$$

where N_t^* denotes the number of modern sector households.

Begin by noting that the first order conditions related to the production decisions in each sector imply labor shares are a constant fraction of output. This allows us to write output in terms of total wages in each sector,

$$\frac{\frac{w_t H_t}{1-\alpha} \;\; + \;\; \tilde{N}_t \frac{\tilde{w}_t f_t}{1-\alpha}}{N_t^* + N_{t+1}^*(T - e_t) + \tilde{N}_t + \tilde{N}_{t+1}(T - e_t)}.$$

Next, note that the effective labor supply of modern sector households, in adult equivalent units, can be written as $h_t(1 - \eta n_{t+1}) + n_{t+1}\gamma\bar{h}(T - e_t) = h_t \dfrac{1}{1+\psi}$. Thus, the net effect of bearing children, even when taking into account child labor, is to reduce the family's adult equivalent work hours from 1 to $\frac{1}{1+\psi}$. The analogous expression for a traditional sector household is $\frac{1}{1+\psi} - \frac{\psi}{1+\psi}\frac{\alpha}{1-\alpha}\frac{f_t}{h_t}$. The effective labor supply from traditional households is smaller because they have more children as a result of the presence of non-labor income later in life. Using this result, the solution for $f_t/\tilde{h}_t$ from (7.9), and noting that $H_t = N_t h_t/(1 + \psi)$, allows us to write worker productivity as

$$\frac{w_t h_t}{1-\alpha}\,\pi_t \frac{\frac{1}{1+\psi} \;\; + \;\; (1-\pi_t)\frac{\tilde{w}_t}{w_t}\frac{1-\alpha}{1-\alpha+\psi}}{1 + (T - e_t)(\pi_t n_{t+1} + (1-\pi_t)\tilde{n}_{t+1})}. \tag{7.14}$$

Now it is easy to see that as π increases, other things constant, output per worker will rise for three reasons. First, workers are more productive per unit of labor supply in the modern sector, as reflected by the wage gap,

$$\frac{w_t}{\tilde{w}_t} = \left[\frac{1+\psi}{1+\psi-\alpha}\right]^{1+\psi} > 1 \;\text{ or }\; \frac{\tilde{w}_t}{w_t} = \left[\frac{1+\psi-\alpha}{1+\psi}\right]^{1+\psi} < 1.$$

Second, households in the modern sector have a higher effective labor supply since they have fewer children,

$$\frac{1}{1+\psi} > \frac{1-\alpha}{1-\alpha+\psi} = \frac{1}{1+\frac{\psi}{1-\alpha}}.$$

Third, there are fewer child workers, as the expression in the denominator

$$\pi_{t+1}n_{t+1} + (1-\pi_{t+1})\tilde{n}_{t+1}$$

shrinks with a rise in π, lowering the supply of labor for a given level of output. This analysis ignores any changes in human capital accumulation and any effects of the structural transformation on saving and physical capital accumulation (to be discussed in Sect. 7.7 and in Chaps. 8 and 9).

In a famous paper, Young (1995) conducts a careful growth accounting for the four Asian Tiger's "Growth Miracles." Hong Kong, Singapore, South Korea, and Taiwan experienced annual growth in labor productivity between 4 and 6% for more than a quarter-century. Hong Kong and Singapore are city states without agricultural sectors and thus did not experience structural transformations in the usual sense. South Korea and Taiwan did see large reallocations of labor away from agriculture and toward manufacturing. Young estimates that the reallocation of labor raised annual growth rates in worker productivity by 0.7 percentage points in South Korea and 0.6 percentage points in Taiwan.

Although increasing growth rates by between ½ and 1% per year is a large effect, it is an underestimate of the impact of the structural transformation. His measurement approach does not link the structural transformation to human and physical capital accumulation through fertility effects. For example, he carefully accounts for many determinants of human capital including a worker's age. The reduction in fertility associated with the structural transformation reduces the fraction of the workforce that is very young and raises human capital per worker. As suggested above there are also links between the structural transformation and saving, as we will see in Chaps. 8 and 9, that raise physical capital per worker.

China has internal restrictions that inhibit the movement of labor from rural to urban areas to a greater degree than in South Korea and Taiwan, as we will discuss in more detail in Chap. 10. These restrictions on labor mobility add to the wage gaps across agriculture and manufacturing sectors that are created by the family farming mechanism highlighted in this chapter. As China has grown, more labor has been allowed to migrate to the manufacturing sector to ease labor shortages. The large gaps in labor productivity across sectors imply a large potential output gain from labor reallocation. Bosworth and Collins (2008) estimate that the reallocation of labor in China raised growth rates in worker productivity by 1.7 percentage points from 1978 to 1993 and by 1.2 percentage points from 1993 to 2004.

7.6.1 Asia and Africa

If several Asian countries have experienced Growth Miracles why haven't more African countries? This is made more puzzling by the fact that one of the most impressive Growth Miracles belongs to Botswana, whose per capita income grew 7.7% per year from 1965 to 1998.[12] Botswana's Growth Miracle arose out of some very unfavorable initial conditions: the country is landlocked (bad for trade), has a tropical climate (bad for health), abundant diamonds (a natural resource curse in other African countries), and who, after gaining independence from Britain in 1966, had little public infrastructure and only two secondary schools in the entire country (Acemoglu et al. 2003). If Botswana can experience a Growth Miracle, why can't other African countries?

McMillan and Rodrik (2011) provide a clue. They document large productivity gaps, between 2 and 4-fold, across traditional and modern sectors of developing countries. Due to these productivity gaps, they view the flow of labor across sectors as an important driver of a country's economic growth. In particular, they find that the labor flows associated with the structural transformation can explain much of the difference in economic growth between Asian and African countries from 1990 to 2005. While Asian countries have generally experienced a flow of labor away from traditional agriculture and informal urban production, enhancing their economic growth due to the labor productivity gap, African labor flows have gone in the opposite direction, lowering their economic growth. From 1990 to 2005, per capita income in Asian countries grew 3 percentage points higher than in Africa. McMillan and Rodrik estimate that labor flows over this period increased Asian growth by 0.54% per year and reduced African growth by 1.3% per year. Thus, 1.84 percentage points of the 3 percentage point difference in growth can be attributed to differences in labor flows.

McMillan and Rodrik explain the different experiences of Asian and Africa based on different responses to increased openness of their economies. In labor abundant Asia, international trade created an expansion in the low-skilled manufacturing sector. In resource abundant Africa, openness created an expansion in the commodity and agricultural sectors. In effect opening their economies to trade has caused Africa to "de-industrialize" causing labor to flow from high to low productivity uses. We first encountered the "De-industrialization" hypothesis in Chap. 6 and will explore it further in Chap. 8. The McMillan and Rodrik study shows that trade can have negative effects on growth by slowing or reversing the structural transformation. Some economists believe these findings support industrial policies

[12] In the last couple of decades, Africa as a whole has experienced above-average growth that has led to some convergence to the living standards of richer countries (Miguel 2009; Young 2012). However, there are concerns over whether the growth can be sustained because it is heavily-dependent on international commodity prices and because progress in the political institutions of many Africa countries is uncertain. In fact, optimism about African convergence has waned in recent years (see, for example, The Economist 2014).

where the government promotes growth by subsidizing manufacturing (see, for example, the discussion in Norman and Stiglitz (2015) and the discussion of South Korea in Chap. 9).

7.7 Productivity Gaps: Measurement and Interpretation

Measuring worker productivity is a difficult task, especially in the traditional sector of developing countries. It is even more difficult to measure the *marginal* product of labor, which is not directly observable. The marginal product of labor is the concept of ultimate interest because gaps in the marginal product tell us about the efficiency of labor allocations. This section discusses these important measurement issues in more detail.

Let's begin with the following expression for the gap in annual worker productivity across sectors,

$$\frac{APL}{A\tilde{P}L} = \frac{APH}{A\tilde{P}H} \times \frac{h}{\tilde{h}} \times \frac{hours}{ho\tilde{u}rs}, \tag{7.15}$$

where APL denotes the average annual product of labor or worker productivity, APH is the average product of a unit of human capital supplied per hour worked, and *hours* refers to the average annual hours worked per worker. In principle, one can use this equation to decompose the sources of the worker productivity gap. The average product of labor, hours worked, and human capital are *relatively* easy to measure. With these measures, the accounting decomposition can be completed by solving for the residual productivity gap that satisfies (7.15), which is interpreted as a gap in the productivity per unit of human capital across sectors.

There are conditions under which the gap in *average* productivity can be interpreted as a gap in *marginal* productivity. In this case, the size of the gap is a measure of inefficiency in the labor market, as an efficient allocation of labor should equate the marginal product of a unit of human capital across sectors. However, before discussing the conditions that imply an equality between average and marginal productivity gaps, we need to recognize that obtaining good estimates of the average product of labor, hours worked, and human capital is a nontrivial task.

7.7.1 Average Product of Labor

To estimate the average product of labor one needs to measure output and the number of workers in each sector. Both are difficult to measure in developing countries because most of the economic activity takes place away from formal markets. In the poorest developing economies, the vast majority of production is consumed at home and not traded. In addition, most workers are family members and sole proprietors, rather than formally paid employees.

The situation is not hopeless because accounting practices have been developed to deal with this problem by focusing on measurements of crop yields rather than on market transactions. In addition, the number of workers, their occupation, and even their hours of work, are collected from household surveys. Gollin et al. (2014) discuss recent measures of production and hours worked in a large data set of 151 countries. They find for the poorest 25% of their sample that the nonagricultural/agricultural productivity gaps average 5.6, with a median value of 4.3. Their measures are roughly consistent with urban/rural consumption gaps, which also rely on household surveys, in developing countries. Young (2013) finds average consumption gaps of around 4.

There is a fair amount of agreement that developing countries are characterized by large gaps in worker productivity in the 4 to 6-fold range. Agreeing on the decomposition and interpretation of these large gaps has proven to be more difficult.

7.7.2 Hours Gap

The proximate causes of the large measured gaps in annual worker productivity can be identified using the accounting decomposition in (7.15). Of the three right-hand-side determinants, hours is the easiest to measure as they are also collected from household surveys. Gollin et al. (2014) measure an annual hours-gap of 1.3 for the poorest 25% of countries in their sample; modern manufacturing workers work 30% more hours than traditional agricultural workers. For some of the poorer countries, the hours-gap is quite large. For example the hours-gap in Uganda is over 2. Their average estimate is in the ballpark of the hours-gap measured for the historical U.S. of 1.4 to 1.5 that we discuss in Chap. 8.

There is not much disagreement over the estimates of the hours-gap. Differences in hours worked explain a significant portion, but not the majority, of the worker productivity gap in developing countries.

7.7.3 Human Capital Gap

Human capital is a more abstract concept than production and hours of work. It must be measured indirectly using methods that have limitations and are open to criticism. Estimates of human capital start by first measuring years of schooling. The years of schooling are translated into a human capital measure by using regression estimates that correlate years of schooling with wages. This common approach to measurement should be viewed as a starting point because the “quality” of a given year of schooling depends on student and teacher attendance throughout the year, teacher qualifications, class size and composition, and other inputs such as books and computers. In addition, it ignores investment in health and in skills obtained from experience on the job.

Using the simplest approach to measuring human capital, based on years of schooling alone, Gollin et al. (2014) find that there is a human capital gap of 1.4

across sectors in the poorest 25% of their country sample. Using their measure of human capital, the combined hours and human capital gaps explain a 1.82 worker productivity gap, leaving an estimated gap in the average product of human capital for these countries of 3 and a median gap of 2.4.

The residual worker productivity gap can be further reduced by attempting to measure the other types of human capital investments mentioned above, which likely differ across sectors as well. One way to begin accounting for the omitted investments is to allow for sector-specific returns to years of schooling. There is evidence that the return to a given year of schooling is lower in agriculture than in nonagricultural. If this is because there are fewer skills associated with *a given year of schooling* acquired by an agricultural worker then agricultural workers have less human capital than the estimates based on schooling alone. This could happen because of *where agricultural workers are educated* (inferior schools or less human capital investment at home in the traditional sector) or because *less able workers self-select into agriculture* (no matter where they are educated). After adjusting for the different rate of return to schooling across sectors, Vollrath (2014) finds little remaining residual gap in a small sample of developing countries. Herrendorf and Schoellman (2014) obtain the same result for the U.S. in current times.

In summary, the residual productivity gap ($APH/\tilde{APH}$) estimated by Gollin, Lagakos, and Waugh of 2 to 3 is quite large. At least some of this gap can be accounted for by adjusting for the different rates of return to schooling across sectors and for other differences in human capital investments that may not be reflected in rates of return to years of schooling (i.e. some of the human capital investments may have a constant or "intercept-effect" that does not alter the estimated return to a marginal year of schooling). It is not clear *why* the returns to schooling are lower in agriculture. The precise reason for the human capital gap has a bearing on whether to interpret the residual productivity gap as a sign of inefficiency. The implied residual estimate of $APH/\tilde{APH}$ from (7.15) is often interpreted as the gap in the *marginal* product of a unit of human capital across sectors, a measure of labor market efficiency. We now turn to assumptions needed to justify this interpretation.

7.7.4 Average and Marginal Products of Human Capital

In the theory we use in this book, the connection between, wages, marginal products, and average products is made clear by the following equalities.

$$\frac{w_t}{\tilde{w}_t} = \frac{(1-\alpha)Ak_t^{\alpha}}{(1-\rho)\tilde{A}(l_t/f_t)^{\rho}} = \frac{APH}{\tilde{APH}}.$$

The first equality says that the rental rate on human capital is equal to the marginal product of human capital in each sector. So, for a given level of human capital, wage ratios equal marginal productivity ratios. However, this is only true under perfectly competitive markets. For example, suppose there was a national market for human capital that forces human capital rental rates to be equated across the two sectors. Suppose further that the two sectors produce different goods and that the modern sector is dominated by a monopolist. In the modern sector, human capital would then *not* be hired until the marginal product is equated to the common rental rate. Profit maximization for a monopolist causes production and employment to fall short of the levels needed to equate the marginal product to the rental rate. This would cause a gap in the marginal products across sectors even when there is no gap in rental rates (or in wages for a given level of human capital). Thus, imperfect competition can cause the first equality to fail.

Next, the exact form of the marginal product expression results from our assumption of Cobb-Douglas technologies. With Cobb-Douglas technologies the marginal product of human capital is equal to the sector labor share times the average product of human capital. If the labor shares are equal across sectors, then the ratio of the marginal products equals the ratio of the average products, the second equality above. However, the equality of the labor shares across sectors is at least somewhat controversial. Vollrath (2009a) and Gollin et al. (2014) argue for equality and Herrendorf and Schoellman (2013) claim the labor shares differ across sectors. To go from the average product estimates directly to the marginal product estimates, requires equality in the factor shares.

7.7.5 The Structural Transformation, Growth, and Economic Efficiency

As discussed in Sect. 7.6, the structural transformation can increase growth in worker productivity by increasing average hours worked, increasing the average age and experience of the work force, and, if the marginal product of human capital is higher in the modern sector, by increasing the productivity of an experience-adjusted hour of work. It is also possible that the structural transformation increases education and human capital. The lower years of schooling observed in the traditional sector may be due to a binding poverty trap that does not exist in the modern sector. This can happen because the relative productivity of children may be lower in the modern sector, either due to more effectively enforced child labor and compulsory schooling laws or because the relative productivity parameter of children (γ) is fundamentally lower in the modern sector due to differences in technologies or types of goods produced. Schooling may also be lower because the price of purchased school inputs, such as quality-adjusted teacher time and school materials, is lower in the modern sector (see the extension of the human capital modeling to include purchased inputs in the problems from Chap. 4).

Even if growth in worker productivity rises due to the structural transformation, it may not indicate a significant improvement in economic efficiency. Suppose, for example, the marginal product of a unit of human capital supplied is roughly equal across sectors, as indicated by the estimates of Vollrath (2014) and Herrendorf and Schoellman (2014). In this case, labor markets are interpreted as efficiently allocating human capital across agriculture and manufacturing. However, it is important to note that efficiency considerations in dual economies should not be restricted to the *allocation* of human capital alone.

We saw in Sect. 7.5 that the lack of efficiently-operating land and insurance markets can bind workers to the traditional sector. This can cause workers to supply less labor hours than they optimally prefer over the year. Also, because intergenerational loan markets are incomplete (see Chaps. 2 and 4), the education of children is dependent on the wages of parents and thus can be lower than is productively efficient (the marginal return to investment in human capital can be greater than the market interest rate). If parental incomes are lower, compulsory schooling laws are less effective, the relative productivity of children is higher, or purchased school inputs are more expensive in the traditional sector, then the underinvestment in years of schooling can be more severe there than in the modern sector. Furthermore, the fact that the marginal returns to schooling, for a given number of years attended, is higher in the modern sector may be due to underinvestment in the health of mothers and young children in the traditional sector, causing the "ability" of traditional sector students to be lower than students in the modern sector. This is a legitimate concern because, generally, public health measures and health outcomes are worse in rural areas than in cities (World Bank 2013).

For all these reasons, the structural transformation can reduce *inefficiencies in hours worked and human capital investments*, even if the economy is allocating a given stock of human capital reasonably well. As rising wages pull workers into the modern sector, fewer workers are dependent on land inheritance and as a result they can freely supply the optimal amount of work effort. Children will be raised by parents with higher wage incomes in environments that lead to more investment in their health and education, reducing the gap between the returns on these investments and the market rate of return on physical assets.

7.8 Conclusion

Persistent wage and productivity gaps across sectors are common features of developing countries. Attempts to explain these gaps while assuming complete markets in the low-productivity traditional sector are not entirely convincing. Institutional arrangements in traditional agricultural and other informal production structures are not well captured by assumptions that markets are complete. One way to distinguish between the two sectors is by the absence of fully developed markets for land and labor in the traditional sector. Regarding land ownership, for example, the evidence suggests that arrangements are made that keep land in the family over many generations. The absence of these factor markets helps to explain the wage gap in two ways.

First, an absence of land markets in the traditional sector fosters the tradition of passing land and farms from one generation of the family to the next. An important condition for receiving an inheritance of this type is that children must remain in the traditional sector to gain the experience needed to operate the farm and "keep it in the family." This institutional arrangement ties occupational choice and geographic location to land ownership in the traditional sector. In equilibrium, traditional workers are willing to take lower wages than those offered in the modern sector because they are compensated by rental income on the land they eventually own and by bigger families.

Second, if the labor market in the traditional sector is also thin, family farms must rely predominantly on family labor. The diminishing returns to work on the family farm lower the reward to work relative to a market setting where the competitive market wage is independent of an individual's labor supply choice. The lower reward to work results in fewer hours worked throughout the year and a larger annual gap in productivity and wages.

In summary, missing markets for land and labor both contribute to large gaps in wages and fertility that can persist for many decades. As a result, the movement of households away from the traditional sector and toward the modern sector increases productive efficiency, reduces population growth, and raises worker productivity.

7.9 Exercises

Questions

1. What are wage and fertility gaps? How large are they? What is their connection to economic growth?
2. Summarize the explanations for wage gaps that do not rely on missing markets. Discuss the weaknesses of each explanation.
3. Offer some reasons why land and labor markets might not exist in the traditional sector. What evidence suggests the absence of land and labor markets in developing countries?
4. Explain why inheritance of the family farm raises fertility. Why does the inheritance have no effect on schooling?
5. Explain how each of the following affect the wage gap.
    (a) traditional sector fertility.
    (b) traditional sector schooling.
    (c) the relative productivity of child labor.
    (d) the strength of parent's preferences for children.
6. Explain the ordering of household consumption across native households in the modern sector, traditional sector migrants to the modern sector, households that remain in the traditional sector. What is the ordering of utility for these household types?
7. Use the model to explain how each of the following change over the course of economic development.
    (a) average fertility in both sectors.
    (b) the wage gap across sectors.
    (c) child labor in both sectors.
    (d) human capital rental rate in the traditional sector.

8. There is an annual wage gap across traditional and modern sectors. Is this due to differences in productivity per hour or differences in hours worked over the course of a year?
9. When would traditional households *choose* to work less than modern households? What is the intuition for this result?
10. Consider the situation when adult work and the quantity and quality of children are all household choice variables. How do schooling, fertility, and work effort interact to determine the wage gap across sectors in this setting?
11. Explain why inefficient or missing markets can reduce labor mobility.
12. Explain how the structural transformation directly raises average worker productivity. How important was the structural transformation in determining the growth rates of the Asian Tiger countries during their Growth Miracle?
13. Discuss the difference in the growth experiences of Asia and Africa and relate the difference to the structural transformation.
14. In today's developing countries how large is the gap in annual labor productivity across sectors? What are the reasons for the gap?
15. How do we know whether human capital is efficiently allocated across sectors? Can the structural transformation cause a rise in average worker productivity even if the developing country allocates units of human capital efficiently across sectors?
16. Explain *G11*.

Problems

1. Derive the following equations and offer an economic explanation for each: (7.7a, 7.7b, 7.7c), (7.8) and (7.11).
2. Assuming that schooling is equal across sectors, write an expression for $\tilde{n}_{t+1}/n_{t+1}$ in terms of the parameters of the model. If $\rho = 0.40$, what value must ψ have if fertility is 50% higher in the traditional sector? Given the required value for ψ, what is the implied wage gap across sectors?
3. If $\alpha = \rho = 1/3$ and $\psi = 1/5$, compute numerical values for each of the following.
    (a) $\tilde{n}_{t+1}/n_{t+1}$
    (b) $w_{t+1}/\tilde{w}_{t+1}$
    (c) the relative supply of effective labor per family in the modern sector
    (d) the relative labor productivity per family in the modern sector
4. Use Vollrath's approach to carefully prove, showing all work, that if there is an absence of labor markets in the traditional sector, then traditional households will choose to work less than modern households.
5. Show that when households choose work, fertility, and schooling of their children, that $\tilde{z}_t < z_t$ and $\tilde{n}_{t+1} > n_{t+1}$, i.e. that traditional households work less and have more children than modern sector households.
6. Show that greater adult human capital raises work effort in (7.12c).
7. Worker *Productivity*. Derive the expression that relates the structural transformation to worker productivity by carrying out the following steps. First, show that output can be written in terms of labor income in both modern and

traditional sectors. Next, write out expressions showing that the total effective labor supply from modern sector households exceed that from traditional households. Finally, write out worker productivity in terms of the wage gap, the household supply of effective labor from each sector, and the fraction of the households in the modern sector, as in (7.14).

8. Suppose an economy is stuck in a poverty trap throughout its structural transformation, with (i) $e_t = \bar{e}$ and (ii) $n_t = 2$ for households that do not inherit land. Using the same parameter values from *Problem* 3 and assuming $T = 0.5$, use (7.14) to compute ratio of worker productivity at the end of the structural transformation, when $\pi_t = 1$, to that at the beginning of the structural transformation, when $\pi_t = 0$.

References

Acemoglu D, Robinson J (2012) Why nations fail. The origins of power, prosperity, and poverty. Crown Publishers, New York.

Acemoglu D, Johnson S, Robinson J (2003) An African success story: Botswana. In: Rodrik D (ed) In search of prosperity. Princeton University Press, Princeton, New Jersey.

Alston L, Hatton T (1991) The earnings gap between agricultural and manufacturing laborers, 1925–1941. Journal of economic history 51(1):83–99.

Au C, Henderson J (2006) How migration restrictions limit agglomeration production in China. Journal of development economics 80(2):350–388.

Banerjee A, Duflo E (2011) Poor economics. Public Affairs, New York.

Basu K (1991) Agrarian structure and economic underdevelopment. Harwood Academic Publishers, New York.

Bertocchi G (2006) The law of primogemiture and the transition from landed aristocracy to industrial democracy. Journal of economic growth 11:43–70.

Bosworth B, Collins S (2008) Accounting for growth: Comparing China and India. Journal of economic perspectives 22(1):45–66.

Calvo G (1978) Urban unemployment and wage determination in LDCs: Trade unions in the Harris-Todaro model. International economics review 19(1):65–81.

Caselli F (2005) Accounting for cross-country income differences. In: Aghion D, Durlauf S (eds) Handbook of economic growth. Elsevier, Amsterdam, pp 679–741.

Caselli F, Coleman W (2001) The US structural transformation and regional convergence: A reinterpretation. Journal of political economy 109(3):584–616.

Chanda A, Dalgaard CJ (2008) Dual economies and international total factor productivity differences: Channeling the impact from institutions, trade, and geography. Economica 75 (300):629–661.

Collier P, Radwan S, Wangwe S (1986) Labour and poverty in rural Tanzania: Ujameen and rural development in the United Republic of Tanzania. New York, Oxford University Press.

Cordoba J, Ripoll M (2006) Agriculture, aggregation, wage caps, and cross-country income differences. Working paper 246, University of Pittsburgh Department of Economics.

David P (1967) The growth of real product in the United States before 1840: New evidence, controlled conjectures. Journal of economic history 27(2):151–197.

David P (2005) Real income and economic welfare growth in the early republic. Stanford University Department of Economics working paper 06-001.

Doepke M (2004) Accounting for the fertility decline during the transition to growth. Journal of economic growth 9(3):347–83.

Drazen A, Eckstein Z (1988) On the organization of rural markets and the process of development. American economic review 78(3):431–43.

Ely R, Wehrwein G (1940) Land economics. Macmillan Co., New York.
Feder G, Just D, Zilberman D (1982) Adoption of agricultural innovation in developing countries: A survey. World Bank staff working paper 542. World Bank, Washington, DC.
Field E (2007) Entitled to work: Urban property rights and labor supply in Peru. Quarterly journal of economics 122(4):1561–1602.
Foster A, Rosenzweig M (1996) Technical change and human capital returns and investments: Evidence from the green revolution. American economic review 86(4):931–53.
French H, Hoyle R (2003) English individualism refuted and asserted: The land market of Earls Coline (Essex), 1550–1750. The economic history review 56(4):595–622.
Galor O (2005) From stagnation to growth: A unified growth theory. In: Aghion P. Durlauf S (eds) Handbook of economic growth. North Holland, Amsterdam.
Galor O, Weil DN (2000) Population, technology, and Growth: From Malthusian stagnation to the demographic transition and beyond. American economic review 90(4):806–828.
Galor O, Moav O, Vollrath D (2009) Inequality in land ownership, the emergence of human capital promoting institutions, and the great divergence. Review of economic studies 76(1):143–179.
Gardner B (2002) American agriculture in the twentieth century. Harvard University Press, Cambridge, Massachusetts.
Goldin C, Katz L (2000) Education and income in the early 20th century: Evidence from the prairies. Journal of economic history 60(3):782–818.
Gollin D, Lagakos D, Waugh M (2014) The agricultural productivity gap. Quarterly journal of economics 129(2):939–993.
Gollin D, Parente S, Rogerson R (2004) Farmwork, homework, and international income differences. Review of economic dynamics 7(4):827–50.
Greenwood J, Seshadri A (2002) The US demographic transition. American economic review. 92 (2):153–59.
Haines M (2000) The population of the United States 1790–1920. In: Engerman S, Gallman R (eds) The Cambridge economic history of the United States. Cambridge University Press, New York.
Hansen G, E Prescott (2002) Malthus to Solow. American economic review 92(4):1205–1218.
Harris B (1972) Innovation adoption in Indian agriculture: The high-yielding variety programs. Modern Asian studies 6.
Harris J, Todaro M (1970) Migration, unemployment, and development: A two-sector analysis. American economic review 60(1):126–42.
Hayashi F, Prescott E (2008) The depressing effect of agricultural institutions on the prewar Japanese economy. Journal of political economy 116(4):573–632.
Hazan M, Berdugo B (2002) Child labor, fertility, and economic growth. Economic journal 112 (482):810–78.
Henderson V (2010) Cities and development. Journal of regional science 50(1):515–540.
Herrendorf B, Schoellman T (2013) Why is measured productivity so low in agriculture? Arizona State University working paper. http://www.public.asu.edu/~tschoell/papers/AgricultureSubmissionNovember2013.pdf.
Herrendorf B, Schoellman T (2014) Wages, human capital, and the allocation of labor across sectors. Arizona State University Working Paper. In: Society for Economic Dynamics 364, 2014 Meeting Papers. https://www.economicdynamics.org/meetpapers/2014/paper_364.pdf
Janvry, A., Emerick, K, Gonzalez-Navarro, M., and Sadoulet, E., 2015, "Delinking Land Rights from Land Use: Certification and Migration in Mexico," *American Economic Review*, 105 (1), 3125–3149.
Jenkins C, Knight J (2002) The economic decline of Zimbabwe. Palgrave Press, New York.
Jerven M (2013) Poor numbers. Cornell University Press, Ithaca, New York.
La Porta R, Shleifer A (2014) Informality and development. Journal of economic perspectives 28(3):109–126.
Lewis A (1954) Development with unlimited supplies of labor. Manchester school 22:139–192.

Lord W, Rangazas P (2006) Fertility and development: The roles of schooling and family production. Journal of economic growth 11(3):229–261.
Macfarlane A (1978a) The origins of English individualism: The family, property, and social transition. Oxford University Press, Oxford.
Macfarlane A (1978b) The origins of English individualism: Some surprises. Theory and society 6(2):255–277.
Mazumdar D (1989) Microeconomic issues of labor markets in developing countries: Analysis and policy implications. EDI Seminar Paper No. 40. World Bank, Washington DC.
McMillan M, Rodrik D (2011) Globalization, structural change, and productivity growth. In: Bacchetta M, Jansen M (eds) Making globalization sustainable. International Labour Organization and World Trade Organization, Geneva.
Miguel E (2009) Africa's turn? MIT Press, Boston.
Moav O (2005) Cheap children and the persistence of poverty. Economic journal 115(500):88–110.
Mourmouras A, Rangazas P (2007) Wage gaps and development: Lessons from U.S. history. IMF working paper 07/105. International Monetary Fund, Washington, DC
Mourmouras A, Rangazas P (2009) Reconciling Kuznets and Habbakuk in a unified growth model. Journal of economic growth 14(2):149–181.
Mourmouras A, Rangazas P. (2013a) Efficient urban bias. Journal of economic geography 13(3):451–471.
Mourmouras A, Rangazas P. (2013b) Wage and fertility gaps in dual economies. Eurasian economic review 3(1):59–83.
Mueller E (1984) The value and allocation of time in Botswana. Journal of development economics 15:329–360.
Munshi K, Rosenzweig M (2009) Why is labor mobility in India so low? Social insurance, inequality, and growth. NBER working paper 14850. National Bureau of Economic Research, Cambridge, Massachusetts.
Munshi K, Rosenzweig M (2016) Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap, American Economic Review 106:46–98
Myint A (2001) International trade and domestic institutional framework. In: Meier G, Stiglitz J (eds) Frontiers of development economics. Oxford University Press, New York.
Norman, A. and Stiglitz, J., 2015, Industrial Policy
Rangazas P (2000) Schooling and growth: A King-Rebelo experiment with human capital. Journal of monetary economics 46(2):397–416.
Rangazas P (2002) The quantity and quality of schooling and U.S. labor productivity growth (1870–2000). Review of economic dynamics 5(4):932–964.
Ranis G, Fei J (1961) A theory of economic development. American economic review 51(4):533–565.
Restuccia D Yang D, Zhu X (2008) Agriculture and aggregate productivity: A quantitative cross-country analysis. Journal of monetary economics 55(2):234–250
Rosenzweig M (1988) Labor markets in low-income countries. In: Chenery HB, Srinivasan TN (eds) Handbook of development economics, vol. 6. Elsevier, Amsterdam.
Schultz T (1964) Transforming traditional agriculture. Yale University Press, New Haven, Connecticut.
Shapiro D, Gebreselassie T (2008) Fertility transitions in sub-Saharan Africa: Falling and stalling. African population studies 23(1):3–23.
Stiglitz J (1974) Wage determination and unemployment in LDCs. Quarterly journal of economics 88(2):194–227.
Stiglitz J (1988) Economic organization, information, and development. In: Chenery HB, Srinivasan TN (eds) Handbook of development economics, vol. 6. Elsevier, Amsterdam.
Temple J (2005) Dual economy models: A primer for growth economists. The Manchester School 73:435–478.
Temple J, Woessmann E (2006) Dualism and cross-country growth regressions. Journal of economic growth 11(3):187–228.

The Economist (2014) Economic convergence: The headwinds return. September 14.
Vollrath D (2009a) How important are dual economy effects for aggregate productivity? Journal of development economics 88(2):325–334.
Vollrath D (2009b) The dual economy in long-run development. Journal of economic growth 14(4):287–312.
Vollrath D (2013) Measuring aggregate agricultural labor effort in dual economies. Eurasian economic review 3(1):39–58.
Vollrath D (2014) The efficiency of human capital allocations in developing countries. University of Houston working paper 201307956.
Wang X, Piesse J (2010) Inequality and the urban-rural divide in China: Effects of regressive taxation. China and the World Economy 18(6):36–55.
Wang X, Piesse J (2011) The micro-foundations of dual economy models. The Manchester School 81: 80–101.
World Bank (1995) African development indicators. World Bank, Washington D.C.
World Bank (2013) Global monitoring report 2013: Rural-urban dynamics and the millennium development goals. World Bank, Washington, DC.
Young A (1992) A tale of two cities: Factor accumulation and technical change in Hong Kong and Singapore. In: Blanchard O, Fisher S (eds) NBER macroeconomic annual. MIT Press for the National Bureau of Economic Research, Cambridge, Massachusetts.
Young A (1995) The tyranny of numbers: Confronting the statistical reality of the east Asian growth experience. Quarterly journal of economics 110(3):641–680.
Young A (2012) The African growth miracle. Journal political economy 120(4):696–739.
Young A (2013) Inequality, the urban-rural gap, and migration. Quarterly journal of economics 128(4):1727–1785.

Physical Capital in Dual Economies

8

In this chapter we shift attention to industrialization and economic growth in dual economies. We focus again on the consequence of missing land markets, a common characteristic of developing economies that was documented in the previous chapter. Here we examine the connections between land ownership, saving, and physical capital formation.

We consider two common situations that are associated with nonexistent or thin land markets. The first situation is one where households that farm the land either own the land or have informal property rights over the income generated from the land. In the case of the U.S., land policy was intentionally designed to transfer ownership of public land holdings across the entire rural population (Libecap 2007). In other cases, the ownership of land remained communal but was effectively divided across local community farmers who were given autonomy in making decisions about land use and had clear claims to income from agricultural production. This occurred in the early stages of development in the Netherlands (de Vries and van de Woulde 1999, Ch. 5), a country where there was no strong feudal legacy, and in African villages (Collier et al. 1986).

The second situation is one where land ownership is highly concentrated in the hands of a relatively few landlords. The landlords do not farm but instead hire landless workers as farm hands. This situation was common in the historical development of Latin America (Deninger and Squire 1997) and Mexico (Acemoglu and Robinson 2012, pp. 35–38). In both of these situations we assume that with thin land markets, the dominant way that land is transferred is through an intergenerational bequest designed to keep land within the family or village.

The inheritance of land during an individual's later years substitutes for saving as a way of financing consumption during old age. This common feature of traditional agriculture undermines the incentive to save when young. Workers that move away from family production to work in factories often lose the claims to family or tribal lands. The loss of a land-bequest implies that workers must save more of their wages to finance retirement consumption. Thus, the structural

S. Das et al., *Economic Growth and Development*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-319-89755-4_8

transformation of the economy away from traditional agriculture and toward modern manufacturing causes an increase in the aggregate saving rate.

When land ownership is concentrated, as in our second scenario, landowners are often able to exert political pressure to change policy in ways that raise land rents. Policies that increase land rents are ones that lower capital accumulation, reduce the opportunities of workers in the modern sector, and keep wages low. Underlying forces that generate a structural transformation, such as technological progress, helps pull workers out of the traditional sector despite anti-growth policies. This weakens the power of land owners and leads to more pro-growth policies which serve to increase capital accumulation and accelerate the structural transformation.

There are other explanations for the rise in capital accumulation over the course of development. Recall from Chap. 1 that Lewis (1954) also thought that a dual economy approach was needed to explain rising saving rates. He conjectured that the income of capital owners in the modern sector would increase relative to incomes of workers and land owners as "surplus" labor from the traditional sector is pulled into the modern sector with little upward pressure on wages. Lewis believed that the relative expansion of capital income was important for growth because capital owners saved a larger fraction of their income than land owners and workers. In contrast to Lewis, our explanation is consistent with rising wages, a constant share of capital income, and an endogenous saving rate that does not depend on the source of income. See Gollin (2014) for a discussion and critique of the classic Lewis model.

Another theory of capital accumulation is based on freeing the entrepreneurial activity of small informal businesses in the traditional sector. The argument is that these entrepreneurs have access to potentially profitable technologies but their investment is held back by credit market imperfections, costly regulations, and high taxes on capital income. While this is likely true in some cases, the evidence suggest that capital accumulation arising out of small informal businesses is small (Banerjee and Duflo 2011, Chap. 9 and La Porta and Shleifer 2014). During development, the vast majority of small informal businesses simply die out as the larger and already established firms operating in the modern sector expand. In our model technological progress, as well as human capital accumulation (see Chap. 9), drives up the productivity of workers in the modern sector causing wages to rise relative to the value of rents from traditional or informal production. The exit of workers from the traditional sector causes informal production to die out and also raises the savings rate, in the manner described above, helping to finance the expansion of capital in the modern sector.

While the absence of markets for land helps motivate the models of this chapter, we assume that there is a competitive labor market, as in the first part of Chap. 7. Labor markets seem to operate more consistently than land markets at early stages of development. After the demise of the European feudal system in the fourteenth century, labor markets quickly formed (Acemoglu and Robinson 2012, Ch. 4). In fact, Clark (2007, Ch. 3) provides evidence of market wages for building and farm laborers that stretch back into the thirteenth century. At least by the nineteenth century, labor markets were working well enough in the U.S. and Europe to virtually eliminate hourly wage gaps across sectors (Mourmouras and Rangazas 2009a, pp. 154–156). Vollrath (2014) argues that labor markets perform well in currently developing countries. He finds only small sector gaps in wages per unit of human capital supplied. However, Gollin et al. (2014) find that large wage gaps still exist after controlling for hours worked and education. While Chap. 7 discusses several explanations for these wage gaps that are consistent with well-working labor markets, their findings suggest that one cannot rule out labor market imperfections at the early stages of development altogether. The approach we take in this chapter is best interpreted as a simplification that allows us to focus solely on the lack of land markets.

Sections 8.1, 8.2, and 8.3 offer three historical applications where the absence of complete land markets was important. The three applications are self-contained and focus on different topics, so readers can pick and choose among the applications based on their interests.

In two of the applications land was owned predominately by small farmers. In Sect. 8.1 we use land inheritance to explain the persistence of large wage gaps in U.S. history. As mentioned above, as well as in Chap. 7, the U.S. wages gaps were not due to differences in hourly wages but rather were caused by differences in the annual hours worked per worker. Annual work hours were much higher in the nonfarm sector, leading to higher annual productivity per worker. Section 8.2 re-introduces the De-industrialization hypothesis from Chap. 6. We focus on the Ottoman Empire in the nineteenth century because it experienced an unusually large increase in the relative price of its agricultural goods as it expanded its trade with Europe. We examine the effects of trade on economic growth and the distribution of income across the population.

Unlike Sects. 8.1 and 8.2, in Sect. 8.3 we assume that landownership is highly concentrated among a few large landowners. As mentioned, concentrated landownership is interesting, in part, because it tends to coincide with political power. We look at the incentives of large landowners to influence fiscal policy in ways that slow the structural transformation. The version of the two-sector model we use in this section also explains a possible connection between the structural transformation and the growth of government as mentioned in *G12*.

8.1 Farmer-Owned Land I—Wages Gap in U.S. History

We maintain the same two sector structure as in Chap. 7, where we have a modern sector of landless households and a traditional sector of landed households, with different production technologies used in each sector. However, now we use the two period overlapping-generations model to allow for saving behavior. Also, to focus attention on physical capital formation, we ignore human capital and assume that fertility is exogenous. Human capital and endogenous fertility will be re-introduced in Chap. 9. Finally, unlike in Chap. 7, we now make explicit the preference to bequeath land to children working on the family farm.[1]

8.1.1 Modern Sector

Producers in the modern sector are neoclassical firms that use the standard Cobb-Douglas production function

$$Y_t = K_t^{\alpha} H_t^{1-\alpha}, \tag{8.1}$$

where Y denotes output, K is the physical capital stock, H is the effective labor supply in the modern sector, and α is the capital share parameter. Given our objectives, it does no harm to reduce notational clutter a bit by assuming $D_t = \tilde{D}_t = A = \tilde{A} \equiv 1$, so we are back to $H = M$.

The firms operate in perfectly competitive markets for goods and inputs. The standard profit-maximizing factor-price equations for the rental rates paid to labor and physical capital are

$$w_t = (1-\alpha)k_t^{\alpha} \tag{8.2a}$$

$$r_t = \alpha k_t^{\alpha-1} \tag{8.2b}$$

where $k = K/M$.

8.1.2 Traditional Sector

The traditional sector differs from the modern sector because of differences in the production technology and because the technology and land used in production is inherited and operated by a family—creating lifetime rental income to supplement wages. For simplicity, we continue to assume that the same goods are produced in each sector. This is a common assumption in the dual economy approach when the specific focus is not on the relative price of the different goods produced in the two sectors. Section 8.2 provides an application where it is important to think of the two sectors as producing distinct goods with a relative price that generally differs from one.

[1]Explicitly introducing the preferences that underlie the desire to bequeath land to children who continue the “farming” tradition does not affect the analysis qualitatively. It is worth showing that it can be done to complete the analysis and ensure that it is logically consistent. After Sect. 8.1 we return to the simpler approach that does not explicitly model the desire to keep land within the family.

Traditional output is produced using the following Cobb-Douglas technology

$$O_t = l_t^{\rho} f_t^{1-\rho}, \tag{8.3}$$

where O output, l is land per farm, f is farm labor per household in the traditional sector, and $0 < \rho < 1$ is a technology parameter. The technology differs from that used in industry because land is an input rather than physical capital. This assumption is meant to capture the idea that production in the traditional sector does not rely heavily on the plant and equipment used in the "factories" of the modern sector. While in general ρ and α may differ, we assume that $\alpha = \rho$. This assumption implies the labor share of income is independent of the stage of development, a stylized fact about economic growth (see, for example, Gollin 2002).

There is no market for the land, either to buy or rent. We assume that there is a market for farm labor, so households could hire labor beyond that provided by the family.

8.1.3 Households

There are two types of households, landed and landless. The landed households inherit the traditional technology/land from their parents. Parents are only willing to pass the land on to those children who remain in the traditional sector full time in order to learn the traditional technology and tend to the farm. The paid labor supply of landed households is constrained by the length of the work-year and by the physical demands of work in the traditional sector. Landless households are free to work in either sector and can freely choose their work time.

All households live for three periods; one period of childhood and two periods of adulthood. In the first period of adulthood, households are endowed with one unit of time. Preferences are identical across households, regardless of occupational choice or residence. All households value their consumption over the two periods of adulthood (c_{1t}, c_{2t+1}), the time spent taking leisure ($1 - z_t$), and the transfer of land to children who continue the tradition of family farming (if land is owned). Preferences are given by

$$U_t = \ln c_{1t} + \beta \ln c_{2t+1} + \xi \ln\left(\kappa + \varphi_{t+1} n b_{t+2}\right) + \zeta \ln(\nu + 1 - z_t), \tag{8.4}$$

where the preference parameters β, ξ, κ, ζ, and ν are nonnegative, where φ gives the fraction of household's children that work in the traditional sector, and where b is the bequest of land passed to each member of the next generation of traditional households. In the absence of a land market, the land inherited by a generation-t household, in the second period of their life, satisfies the condition

$b_{t+1} = l_{t+1} = \varphi_{t+1} n b_{t+2}$, i.e. all land that is inherited is ultimately bequeathed to the next generation.[2]

The preference specification includes a term that involves the bequest of land made to the children who continue the tradition of family farming. However, while households value land bequests, it is not essential if $\kappa > 0$. This means that traditional households may be willing to give up their claim to family land to move to the modern sector under the right conditions. In similar fashion, leisure is valued but is not essential—it may be optimal for households to take no leisure if $\nu > 0$.

1. *Landless Households*

Landless households choose what fraction (z_t) of time they devote to work and what fraction to leisure. The landless household maximizes utility subject to the lifetime budget constraint,

$$c_{1t} + \frac{c_{2t+1}}{R_t} = w_t z_t$$

where the return on the ownership of physical capital, purchased in period t and rented to producers in period $t+1$, is $R_t \equiv r_{t+1} + 1 - \delta$. To simplify some of the expressions, we assume $\delta = 1$, so that $R_t = r_{t+1}$.

In addition to the standard first order conditions for life-cycle consumption, the choice of z_t yields

$$\frac{\zeta}{\nu + 1 - z_t} = \lambda_t w_t \tag{8.5}$$

where λ_t is the Lagrange multiplier. Equation (8.5) says that the marginal value of taking more leisure must equal the marginal value of the forgone consumption associated with working less.

Solving the model gives us the following saving and work behavior

$$s_t = \left[\frac{\beta}{1+\beta}\right] w_t z_t \tag{8.6a}$$

$$z_t = \frac{(\nu+1)(1+\beta)}{1+\beta+\zeta} \tag{8.6b}$$

As is clear from (8.6b), the optimal time devoted to work is constant. To simplify notation and reduce the number of parameters, we set $z_t = 1$ by setting $\zeta = \nu(1+\beta)$. This means that landless households provide one unit of labor when young, as has been assumed in previous chapters.

[2] We assume that the land is passed at death, rather than inter vivos, because this binds the children expecting to receive land to work on the family farm in their first period of adulthood.

2. *Landed Households*

In the case of landed households, we assume that the skills and land needed to operate a traditional technology are inherited from parents, but only if children remain in the traditional sector when young. Time spent producing in the traditional sector is exogenously constrained so that $\tilde{z} < 1$. Thus, households that reside in the traditional sector are only able to work for a fraction $\tilde{z}$ of the time that modern sector households do.

The motivation for this assumption is that traditional production involving land is constrained by the weather and the length of day, both of which vary with the seasons. In addition, some types of traditional sector work are too physically strenuous for women and younger children, limiting their labor input relative to that of women and children in the modern sector.[3] In equilibrium, it will be the case that the constraint on the work of traditional workers is compensated for by the land rents they receive from inheriting land and the traditional technology.[4]

Operating the traditional technology generates land rents for the household during the second period of adulthood. The lifetime budget constraint for households choosing to follow their parents and work in the traditional sector is then

$$\tilde{c}_{1t} + \frac{\tilde{c}_{2t+1}}{R_t} = w_t\tilde{z} + \frac{O_{t+1} - w_{t+1}f_{t+1}}{R_t}. \tag{8.7}$$

The timing of the land bequest, and the fact that residual income accrues in the second period of adulthood, captures the idea that the family farm provides a source of retirement income that substitutes for retirement saving. The presence of land as a source of retirement income is a potentially important factor in slowing physical capital accumulation during the early stages of development (see Carter et al. 2003).

Also note that traditional workers are paid the same per hour as modern sector workers. Remember that landless workers can work in either sector. If the traditional sector employs some landless workers, then the wages in each sector must be the same. Landed workers will therefore make the same *hourly* wage as landless workers but make a lower *annual* wage because they work fewer hours over the course of the year.

Maximizing utility subject to (8.4) and (8.7) yields the following equations for traditional household saving and for the farm labor demanded by a traditional producer

[3]The physical demands of farming help explain why the onset of the cottage industries and early factories increased the employment of women and children in the early stages of the Industrial Revolution, especially in regions dominated by the more strenuous Northern farming—hay, wheat, and dairy farming. For evidence in England, see Cunningham (1990) and Horrell and Humphries (1995). For the U.S., see Goldin and Sokoloff (1984).

[4]Note that because preferences are identical, households in the traditional sector would prefer to work as much as those in the modern sector.

$$\tilde{s}_t = \left[\beta - \frac{\rho O_{t+1}}{R_t w_t \tilde{z}}\right] \frac{w_t \tilde{z}}{1+\beta} \tag{8.8a}$$

$$f_{t+1} = \left[\frac{1-\alpha}{w_{t+1}}\right]^{\frac{1}{\alpha}} l_{t+1}. \tag{8.8b}$$

The savings function in (8.8a) exhibits the effect of family farming suggested by Carter et al. (2003). The presence of second period income from operating the traditional technology discourages retirement saving. The fraction of current period wages that is saved is smaller as a result. Thus, saving per household is lower in the traditional sector for two reasons: wages are lower, because of the shorter paid work-year, and there is a smaller fraction of wages that is saved. Equation (8.8b) gives the labor demand by a traditional sector producer similar to that from Chap. 7.

8.1.4 Equilibrium

We now determine equilibrium in the factor markets for labor and physical capital. Given k_t and l_t, the market clearing values of the factor prices and the labor demand chosen by a traditional sector producer are given by (8.2a), (8.2b), and the period-t version of (8.8b). This leaves only the equilibrium values of k_t and l_t to be determined.

1. *Equilibrium* l_t

Determining the equilibrium quantity of land per traditional producer is equivalent to finding the number of traditional sector producers in equilibrium (since total land is fixed). For a household to remain in the traditional sector and inherit land from their parents, the lifetime utility of staying home must be at least as great as the utility associated with forgoing the inheritance and becoming unrestricted in the choice of occupation.

We assume throughout that land is sufficiently scarce to cause some children born in the traditional sector to choose work in the modern sector. In other words, if all children born in the traditional sector remained in the traditional sector as adults, the quantity of land per producer would be too small to generate sufficient land rents to compensate for the constraint on paid work, causing the lifetime utility of traditional sector households to fall below that of landless households. The "competition" for land among traditional sector siblings must be strong enough to create an equilibrium exit from the traditional sector that equates the lifetime utility of the landed and landless households in each generation. This scenario matches what we see in history, a gradual structural transformation away from traditional production to modern production.

The fact that the lifetime utility of the landed and landless households must be equated, gives an equilibrium condition for determining, along with (8.2), the number of traditional sector producers in each period. Equating the indirect utility functions for landed and landless household implies that land rents must satisfy

$$\frac{\alpha O_{t+1}}{R_t w_t \tilde{z}} = \Omega_{t+1} \tag{8.9}$$

$$\text{where } \Omega_{t+1} \equiv \left(\frac{1}{\tilde{z}}\right)\left(\frac{\nu}{\nu + 1 - \tilde{z}}\right)^{\nu}\left(\frac{\kappa}{\kappa + l_{t+1}}\right)^{\frac{\xi}{1+\beta}} - 1.$$

Equation (8.9) states that the equilibrium ratio of the present value of land rents to potential wages in the traditional sector is affected by the extent of the labor-supply constraint in the traditional sector. One can show that the more severe the constraint, i.e. the lower is $\tilde{z}$, the higher is the required land-rent ratio. Actual land rents per household vary inversely with the number of traditional sector households. Thus, the more severe the constraint, the smaller is the total number of households that are able to produce in the traditional sector in order to generate sufficient land rents for those that stay.

The size of land holding per household, and their value as a bequest, also affects the equilibrium land-rent ratio. The larger are land holdings, and the more bequests to children are valued, the lower is the equilibrium land-ratio. Intuitively, the more direct utility that is generated from owning and bequeathing land, the less income compensation is needed to offset the labor supply constraint.

2. *A National Labor Market*

Before solving for the equilibrium capital intensity, we need to establish how much effective labor will be allocated to each sector.

We know that the total effective labor supply for the entire economy in period t is $\pi_t N_t + (1 - \pi_t) N_t \tilde{z}$, where π denotes the fraction of young households that reside in the modern sector, as in Chap. 6.[5] The labor allocated to the traditional sector in period t must equal the traditional producer's demand for farm labor multiplied times the number of farm owners. Farm owners in period t are the older households that remained in the traditional sector to live and work in period $t-1$. The total number of older traditional households that own farms in period t is therefore $(1 - \pi_{t-1})N_{t-1}$. The total demand for labor in the traditional sector is then $f_t(1 - \pi_{t-1})N_{t-1}$. This implies that the effective labor supply in the modern sector is simply the remainder of the work force,

[5]It is trickier to keep track of households in this chapter than in Chap. 6. In Chap. 6 it did not actually matter where we thought of workers residing because all households could work in either sector. Here, landless households can work in either sector, but landed households must reside and work in the traditional sector.

$$M_t = \pi_t N_t + (1 - \pi_t) N_t \tilde{z} - f_t (1 - \pi_{t-1}) N_{t-1}. \tag{8.10}$$

An additional equilibrium condition is needed to ensure that the traditional sector is not constrained in the labor its uses, i.e. the demand for labor from the traditional sector must be greater than or equal to the supply of labor coming from landed households. If this were not the case, then there would be a "dual" labor market that could give rise to different wages per hour across sectors and across household types. Some hiring of part-time labor from landless households, "farm hands," is necessary to integrate the labor markets. Thus, an integrated labor market must satisfy

$$f_t (1 - \pi_{t-1}) N_{t-1} \geq (1 - \pi_t) N_t \tilde{z}, \tag{8.11}$$

the total demand for labor by traditional sector landowners must exceed the labor that is collectively provided by their children that decide to work and remain in the traditional sector.

3. *Equilibrium* k_t

Physical capital intensity is determined by the equilibrium condition that equates the physical capital demanded for production in the modern sector to the supply of capital that results from retirement savings of the households one period before,

$$K_{t+1} = \pi_t N_t s_t + (1 - \pi_t) N_t \tilde{s}_t. \tag{8.12}$$

Dividing both sides of (8.12) by M_{t+1}, defining $k_{t+1} \equiv K_{t+1}/M_{t+1}$, then using (8.9) and (8.10) allows us to write the transition equation for capital intensity as

$$k_{t+1} = \frac{[\pi_t \beta + (1 - \pi_t)(\beta - \Omega_{t+1}) \tilde{z}](1 - \alpha) k_t^{\alpha} + (1 + \beta)(L/N_t)}{(1 + \beta) n (\pi_{t+1} + (1 - \pi_{t+1}) \tilde{z})}. \tag{8.13}$$

We can now discuss the avenues through which the structural transformation from traditional to modern production affects physical capital intensity. First, the movement of households out of the traditional sector increases saving because it increases wage income ($\tilde{z} < 1$) and because it increases the rate of saving ($\beta - \Omega_{t+1} < \beta$).

The positive effect on saving is countered by a negative effect that occurs as the supply of workers that must be absorbed by the modern sector rises, causing the effective work force that must be supplied with capital to rise, thereby lowering physical capital intensity. Equation (8.10), applied to period $t + 1$, and (8.8b) can be used to write out the labor that must be absorbed by the modern sector as $\left[\pi_{t+1} + (1 - \pi_{t+1}) \tilde{z} - \frac{1}{k_{t+1}} \frac{L}{n N_t}\right] N_{t+1}$. More households living in the modern sector (higher π_{t+1}) and a weaker demand for farm labor because of less land available per old household (lower L/N_t), both cause the labor that must be supplied with capital to rise and k_{t+1} to fall. This explains the appearance of $\pi_{t+1} + (1 - \pi_{t+1})\tilde{z}$ in the denominator and L/N_t in the numerator of the solution for k_{t+1}.

In summary, the overall effect of the structural transformation on physical capital intensity is ambiguous. The positive effects stemming from an increased saving rate must be weighed against the negative effect of needing to provide capital to a growing modern sector work force. Qualitatively, the transition equation, and the economic forces at work, is similar to the transition equation from Chap. 6, where we assumed perfectly competitive land markets (see Eq. (8.11) from Chap. 6). There the relative size of land holdings, as an alternative mode of saving, and the fraction of the work force in the modern sector both served to slow the growth in capital intensity. The same forces are at work here. The relative size of land holdings lower saving and capital intensity via Ω_{t+1}, the equilibrium ratio of land rents relative to wages.

The main difference for capital accumulation in chapter is, *without* a land market, the *traditional* households inherit land and save less in the form of physical capital (and may even borrow). *With* land markets, the saving of *all* households are equally diverted away from physical capital when land is purchased. The relatively low saving of traditional sector households creates a positive effect on capital accumulation associated with the structural transformation that was not present when there were perfectly competitive land markets in Chap. 6. Now, the saving rate of the economy will rise as households move from the traditional to the modern sector because of the different saving behavior of the households in the two sectors.

8.1.5 Labor Productivity

To examine how the structural transformation affects economic growth, we need to examine the behavior of two different measures of worker productivity. To begin, total output of the economy, $Y_t + O_t(1 - \pi_{t-1})N_{t-1}$, can be written in terms of k_t and π_t as

$$k_t^{\alpha}[\pi_t + (1 - \pi_t)\tilde{z}]N_t. \tag{8.14}$$

To compute labor productivity it is common to divide total output by either the number of *hours worked* ($[\pi_t + (1 - \pi_t)\tilde{z}]N_t$) or the number of *workers* (N_t). Labor productivity per hour worked is simply k_t^{α}, while labor productivity per worker is $k_t^{\alpha}[\pi_t + (1 - \pi_t)\tilde{z}]$. The structural transformation can only affect labor productivity per hour indirectly by altering the economy's capital intensity. However, the structural transformation has a direct effect on labor productivity per worker because as labor moves from the traditional to the modern sector, hours worked increase. This distinction can be quite important, as in the case of U.S. economic growth in the nineteenth century.

Table 8.1, taken from Mourmouras and Rangazas (2009a), presents both the relative annual output per worker and the annual wage in nonfarm occupations relative to those in farming. While not as large as those observed in some developing countries today, the wage and productivity gaps in U.S. history were substantial.

Table 8.1 Ratio of nonfarm to farm productivity/wages in the U.S

Time period	Output per worker	Annual nominal wages	Annual real wages
1840	2.0–2.5[a]		
1880	2.7[b]	5.0[b]	2.7[d]
1940	2.5[b]	2.9[b]; 3.3[c]	1.8[c]

Notes: Values are the ratio of annual value for the nonfarm sector to the annual value to the farm sector
[a]David (1967), [b]Caselli and Coleman (2001), [c]Alston and Hatton (1991), and [d]Ratio assuming same correction as Alston and Hatton for 1940

Table 8.2 Hours worked (full time equivalents)

Time period	Nonfarm	Farm
1840[a]	67.8	
1880[b]	62.7	45.5
1940[b]	42.2	44.9

Notes: [a]Margo (2000) and [b]Kendrick (1961, Table A-IX)

Note that the nominal annual wage ratios measured by Caselli and Coleman (2001) are much larger than the productivity ratios. There are at least two reasons for this. First, the wage ratios do not include non-cash compensation, a relatively large portion of the compensation paid to farm workers. Second, there is no adjustment for the fact that farm workers lived in rural areas where the cost of living was lower. Alston and Hatton (1991) show that accounting for these two factors is important. They find that in 1940 monthly cash earnings in manufacturing were 3.3 times those in agriculture, even greater than the Caselli and Coleman estimate of a 2.9-fold wage gap for that year. However, after adjusting for non-cash payments and cost-of-living differences, Alston and Hatton find that the gap shrinks to a 1.8-fold difference. If we apply the same proportional adjustment to the nominal wage in 1880 we get a real wage ratio of 2.7. Thus, after adjustment, the real wage and productivity gaps are roughly consistent and are in the 2–3 fold range during the nineteenth and early twentieth centuries.

The gaps in annual wages and productivity per worker could be due to differences in hours worked per worker or differences in productivity per hour. The evidence suggests that the vast majority of the gap is due to differences in hours worked per year. Table 8.2 documents the fact that hours worked by fulltime employees were higher in the nonfarm sector. David (1967) uses data from Kendrick (1961) to argue that the hours gap was larger than is indicated by the difference in hours worked by fulltime employees because more of the labor in agriculture is part-time (including family members). David estimates that, over the course of a year, the average agricultural laborer worked about half the hours of the average laborer in industry in 1840. Thus, the difference in hours worked alone could explain a twofold gap in annual productivity.

The literature attempting to measure the historical gaps in wages earned per hour are consistent with David's estimate. The main finding in this literature is that the

Table 8.3 Growth rates in U.S. history

Time period	Per hour worked	Per worker
1820	0.39	0.31
1840	0.39	1.82
1860	0.56	1.32
1880	1.06	1.84
1900	1.53	1.53

gaps in wages per hour across industry and agriculture were very small. Margo (2000) estimates very small wage gaps of 10–15% for the United States in the nineteenth century. Hatton and Williamson (1992) find relatively small wage gaps in the United States from 1890 to 1920, ranging from 5% to 30%. They find larger gaps from 1920 to 1940 which are mostly explained by the high urban unemployment rates during the Great Depression.

In summary, the data above indicates that there were 2–3 fold gaps in the annual productivity and annual wages across nonfarm and farm sectors in the U.S. during the nineteenth and early twentieth centuries. The data also suggests that the wage gap was primarily due to the greater number of hours worked per worker in the nonfarm sector. This assessment of the productivity/wage gaps implies there should be a significant difference between the growth in labor productivity per *worker* and per *hour-worked* during the structural transformation.

When there are significant differences in worker productivity per *hour-worked*, then as workers move from agriculture to industry, there should be a significant increase in the average productivity of an hour worked in the economy as a whole. On the other hand, if the sector-productivity gap is largely due to differences in hours spent producing measured output, then as labor migrates to industry, the growth in productivity per hour-worked would not show much rise. However, measured productivity *per worker* would rise significantly due to an increase in hours-worked per worker.

Table 8.3 presents the U.S. economy's growth rate in output per worker and the growth rate in output per hour-worked. During the nineteenth century, there were large differences in the two growth rates. The annualized growth rate for output per worker was 1.36% over the nineteenth century compared to 0.65% for output per hour-worked. An average growth rate in output per worker double that of output per hour-worked supports the claim that the main cause for lower output per worker in agriculture was fewer hours worked. As farm workers migrated to industry over the century, output per worker expanded significantly due to an increase in hours worked per worker. Output per hour worked did not increase as much because farm workers hourly productivity was not much different from that of nonfarm labor.

For output per hour worked to grow only weakly in the early stages of growth there cannot be large changes in physical capital intensity. This means the conflicting effects of the structural transformation on k discussed earlier in this section must approximately offset. Mourmouras and Rangazas (2009a) show that this was the case in the U.S. during the nineteenth century. Throughout most of the nineteenth century, growth in productivity per hour worked was due to technological change and not private capital accumulation per worker.

By the end of the nineteenth century and early twentieth century the growth rates equalized. The change in the relationship between the two growth rates is likely explained by two common features of development. First, the end of the structural transformation saw the size of the agricultural sector decline to a very small and stable fraction of the labor force, thereby ending the movement of labor from a low to a high productivity sector and slowing the growth rate of output per worker relative to output per hour worked. Second, the secular decline in hours worked off the farm further slowed the growth in output per worker relative to output per hour worked. In the U.S. the decline in average hours worked off the farm was significant, as indicated on Table 8.2.

8.2 Farmer-Owned Land II—De-industrialization in the Ottoman Empire*

From Chap. 6, we know that one explanation for the beginning of the Great Divergence focuses on international trade between the rapidly industrializing countries of Western Europe, particularly Great Britain, and a "periphery" of slowly industrializing countries (see, for example, Williamson 2011). Countries in the periphery had a comparative advantage in making primary products. Expanding international trade in the nineteenth century caused the relative price of primary products to increase in the periphery and decrease in Western Europe. As trade expanded, the change in relative prices caused de-industrialization in the periphery and accelerated industrialization in Western Europe. According to the theory, de-industrialization slowed economic growth in the periphery and sped economic growth in Western Europe, thus contributing to the Great Divergence.

We examine the de-industrialization hypothesis as it applies to the nineteenth century Ottoman Empire. The Empire experienced an unusually large increase in its relative price. The relative price of primary products rose 2–2.5 fold in the middle of the century (Pamuck and Williamson 2011). What impact did the dramatic rise in relative prices have on industrialization and growth? Would the standard of living have been higher in the early Turkish Republic of the twentieth century had the rulers of the Ottoman Empire not opened trade with Western Europe at the beginning of the nineteenth century and avoided this rise in the relative price of primary products?

We answer these questions by constructing a dynamic version of the Specific Factor Model that captures some of the key features and stylized facts of the Ottoman Empire.[6] The purpose of the exercise is to imagine what the Ottoman Empire might have looked like if there had not been a dramatic rise in the international price of

[6]Static versions of the Specific Factors model date back to Jones (1971) and Samuleson (1971). A dynamic version of the model was first presented by Eaton (1987).

primary commodities in the middle of the century. In other words, we examine the quantitative significance of De-Industrialization hypothesis for the Empire.

There are two sectors of production, agriculture and manufacturing. The economy is open and the relative price of agricultural goods is determined in international markets. In the agricultural sector, land and labor are used to produce food and primary products. Older generations of farmers have informal ownership claims on the land. The adult children of the farmers, along with migrant workers, provide the farm labor. There is no active market for land. The right to farm the land is passed from the older generation to the younger generation of farmers as a bequest. In the manufacturing sector, skilled craftsmen work in informal shops. They rent physical capital and hire unskilled migrant workers to assist them in production. The relative size of the rural and urban population is determined by where the migrant workers are employed. In summary, the model is essentially a dynamic version of the Specific Factors model that is used to study international trade in developing economies. Here the specific factors are land in agriculture, and capital and skilled craftsmen in manufacturing.

We calibrate the model to be consistent with the following stylized facts related to the Ottoman economy in the period from 1820 to 1920.[7]

1. Weakly positive growth in per capita real income (less than 1%).
2. Positive population growth (approximately 1%).
3. Increasing real wage in manufacturing units, but falling real wages in agricultural units during mid-century.
4. Modest urbanization (from 17% to 22%).
5. Dramatic increase in the relative price of primary products in mid-century (a 2–2.5-fold rise).
6. High interest rates throughout the century (between 10% and 30% annually).

We then compare aggregate growth and the welfare of different groups within the economy had there been *no* increase in the international price of primary products.

8.2.1 The Model

The model is an extension of the standard two-period overlapping-generations model of physical capital accumulation. Young households provide an exogenous unit of labor to production during the first or working period of their life. The wages they earn are used for consumption and savings in the form of physical capital purchases. In the second period, all households retire. They finance retirement consumption from the physical capital they own and rent to manufacturing firms

[7]For more details see Mourmouras and Rangazas (2014).

in exchange for a rental fee. Farmers have a second source of income for retirement consumption in the form of land rents.

The preferences of all households are identical. Lifetime utility for a household of generation-t, takes the form

$$U_t = u_{1t} + \beta u_{2t+1}$$

where β is the time discount factor that values utility from future consumption relative to current consumption. The single period utility flows result from the consumption of both manufacturing and agricultural goods,

$$u_{it} = \chi \ln y_{it} + (1 - \chi) \ln o_{it}$$

where y is the manufacturing good and o is the agricultural good, and χ is a preference parameter that weighs the utility from consuming the two different goods.

Within each age-cohort, there are three different household types: unskilled migrant workers that can be used to produce both goods, skilled craftsmen that produce manufacturing goods, and farmer/landowners that produce agricultural goods.

8.2.2 Migrants

Migrant workers are mobile across sectors. They provide unskilled labor that is used in the production of both types of goods, so there is an aggregate labor market for their services. They are free to work and reside in either the rural area, where they work as farm hands, or in urban areas, where they work in the shops assisting craftsman in making manufacturing goods. They receive the same competitive market wage in either location.

The two single-period budget constraints of migrant workers take the form,

$$y_{1t} + p_t o_{1t} + s_t = w_t D_t \tag{8.15a}$$

$$y_{2t+1} + p_{t+1} o_{2t+1} = R_t s_t, \tag{8.15b}$$

where p is the internationally determined relative price of agricultural goods, s is saving via the purchases of physical capital, w is the competitive wage per unit of effective labor supply in units of the manufacturing goods, D is the exogenous labor productivity index that determines the effective labor supply associated with one unit of labor time, and $R_t \equiv 1 + r_{t+1} - \delta$ is the total return to one unit of physical capital. For simplicity, we assume that capital fully depreciates when used in production, so the rate of depreciation (δ) equals one. This implies the return to capital is completely determined by the competitive rental rate paid to use capital in manufacturing production (r).

Note that we assume that the same labor productivity index applies regardless of where the worker is employed. If the productivity index increases, it is due to *balanced* technological change across the sectors. The technological progress could be occurring for different reasons in the two sectors (e.g. expanding cultivation of raw land in agriculture and new types of physical capital used in manufacturing), but there is no systematic difference in the resulting exogenous growth in worker productivity across sectors. The economy-wide growth rate in D is denoted by d.

The optimal behavior of the migrant workers is given by the following functions: $y_{1t} = \frac{\chi}{1+\beta} w_t D_t$, $p_t o_{1t} = \frac{1-\chi}{1+\beta} w_t D_t$, $y_{2t+1} = \frac{\chi}{1+\beta} \beta R_t w_t D_t$, $p_{t+1} o_{2t+1} = \frac{1-\chi}{1+\beta} \beta R_t w_t D_t$, and $s_t = \frac{\beta}{1+\beta} w_t D_t$. These demand functions simply say that present value of expenditures on goods and assets are a function of wages in manufacturing units.

8.2.3 Craftsmen

Craftsmen individually operate a small manufacturing firm or "shop." One can interpret the craftsman as being truly skilled in the sense that they are uniquely able to operate the shop technology. Alternatively, they may be interpreted as being no different than the unskilled labor and are simply the lucky recipients of government licenses that allow them to operate local monopolies in the urban area. They supply one unit of untraded labor for which they are compensated with a residual rent that can be interpreted either as an implicit skilled wage or as monopoly profit. In addition, the craftsman rent physical capital (k) and hire unskilled migrant labor (m) in competitive factor markets. The craftsman's production technology is given by a Cobb-Douglas production function in capital, unskilled labor, and skilled labor.

$$y_t = k_t^{\alpha} (D_t m_t)^{\overline{\alpha}} D_t^{1-\overline{\alpha}-\alpha}, \tag{8.16}$$

where $\overline{\alpha}$ is the output share of the unskilled migrant labor and $1 - \overline{\alpha} - \alpha$ is the output share of the skilled craftsman. Labor-augmenting technological progress increases the productivity of both skilled and unskilled labor.

The craftsman chooses k and m to maximize profit, resulting in the following first order conditions, that equate the marginal products of these inputs to their competitive factor prices,

$$\alpha \widehat{k}_t^{\alpha-1} m_{t\overline{\alpha}} = r_t \tag{8.17a}$$

$$\overline{\alpha} \widehat{k}_t^{\alpha} m_t^{\overline{\alpha}-1} = w_t, \tag{8.17b}$$

where $\widehat{k}$ is k/D, the "detrended" capital stock. This measure of capital intensity is per craftsman or per urban shop. It differs from the measure of capital intensity in Sect. 6.1, which is per modern sector worker. It is not natural to define capital intensity per modern sector worker here because $1 - \alpha \neq \overline{\alpha}$ and the factor price equations do not simplify in the usual manner by combining k and m. Note also that we can write the profit, or income paid to the skilled craftsman, as $D_t w_t^* = D_t\left(1-\alpha-\overline{\alpha}\right)\widehat{k}_t^{\alpha} m_{t\overline{\alpha}}$, so that the skill premium is

$$\frac{\frac{w_t^*}{w_t} = 1-\alpha-\overline{\alpha}}{\overline{\alpha}} m_t. \tag{8.17c}$$

Except for the difference in wages, the demand functions of the craftsman are perfectly analogous to those of the unskilled laborer: $y_{1t}^* = \frac{\chi}{1+\beta} w_t^* D_t$, $p_t o_{1t}^* = \frac{1-\chi}{1+\beta} w_t^* D_t$, $y_{2t+1}^* = \frac{\chi}{1+\beta} \beta R_t w_t^* D_t$, $p_{t+1} o_{2t+1}^* = \frac{1-\chi}{1+\beta} \beta R_t w_t^* D_t$, and $s_t^* = \frac{\beta}{1+\beta} w_t^* D_t$.

8.2.4 Farmers

The farm technology used to produce agricultural goods is

$$O_t = l_t^{\alpha}(D_t f_t)^{1-\alpha}, \tag{8.18}$$

where l is the land holdings of an individual farm and f is farm labor. Farm labor in period t is unskilled and consists of the young, generation-t, members of the farm family that have informal claims to the use of land, as well as migrant workers that are landless and hired as farm hands. Both types of labor are paid the unskilled wage, $w_t D_t$.

The rights to farm the land are passed down to generation-t members of the farm family when they become old in period $t+1$. The old members of the farm family manage the farm and generate residual income, $p_{t+1} O_{t+1} - w_{t+1} D_{t+1} f_{t+1}$, to help finance retirement consumption. Note that if there is positive population growth, the *individual* farm managed by each of the old households becomes smaller than those managed by the previous generation, because the land holdings are split across all old family members. This is one reason why there is migration to the city.

To maximize residual income, or land rents, the amount of farm labor hired satisfies

$$P_{t+1}(1-\alpha)\left(D_{t+1} f_{t+1}\right)^{-\alpha} l_{t+1}^{\alpha} = w_{t+1}. \tag{8.19}$$

Using (8.24), we can define rental income from farming as $rent_{t+1} \equiv \alpha p_{t+1} O_{t+1} = \frac{\alpha}{1-\alpha} w_{t+1} D_{t+1} f_{t+1} = \alpha p_{t+1}^{1/\alpha}\left(\frac{1-\alpha}{w_{t+1}}\right)^{\frac{1-\alpha}{\alpha}} l_{t+1}$. The present value of the lifetime income of a farm household can then be written as, $w_t D_t\left[1+\frac{rent_{t+1}}{w_t D_t R_t}\right]$. The demand functions for a farm household are

$$\tilde{y}_{1t} = \frac{\chi}{1+\beta} w_t D_t \left[1 + \frac{rent_{t+1}}{w_t D_t R_t}\right], \;\; p_t \tilde{o}_{1t} = \frac{1-\chi}{1+\beta} w_t D_t \left[1 + \frac{rent_{t+1}}{w_t D_t R_t}\right],$$
$$\tilde{y}_{2t+1} = \frac{\chi}{1+\beta} \beta R_t w_t D_t \left[1 + \frac{rent_{t+1}}{w_t D_t R_t}\right], \;\; p_{t+1} \tilde{o}_{1t+1} = \frac{1-\chi}{1+\beta} \beta R_t w_t D_t \left[1 + \frac{rent_{t+1}}{w_t D_t R_t}\right],$$
$$\text{and } \tilde{s}_t = w_t D_t \left[\frac{\beta}{1+\beta} - \frac{1}{1+\beta} \frac{\alpha}{1-\alpha} \frac{rent_{t+1}}{w_t D_t R_t}\right].$$

Note especially that land rent substitutes for retirement income from capital and thus reduces saving.

8.2.5 Labor Market Equilibrium

The number of young households of our three types are denoted as $\bar{N}_t$ (migrant workers), N_t^* (craftsmen), and $\tilde{N}_t$ (farmers). Labor market equilibrium requires that the demand for workers by craftsmen and farmers (net of young members of the farm households) must equal the supply of migrant workers looking for work,

$$N_t^* m_t + \tilde{N}_{t-1} f_t - \tilde{N}_t = \bar{N}_t. \tag{8.20}$$

Using the fact that $l_t = L/\tilde{N}_{t-1}$, where L is the total amount of raw land available for farming, and (8.17b) and (8.19) to solve for the demand for labor by craftsmen and farmers, allows us to rewrite (8.20) as

$$N_t^* \left(\frac{\overline{\alpha}}{w_t} \widehat{k}_t^{\alpha}\right)^{\frac{1}{1-\overline{\alpha}}} + \left(\frac{(1-\alpha)p_t}{w_t}\right)^{\frac{1}{\alpha}} \left(\frac{L}{D_t}\right) = \bar{N}_t + \tilde{N}_t. \tag{8.21}$$

Note that technological progress reduces the demand for labor by farmers because, with greater productivity per worker, fewer farm hands are needed to work the available land. This inverse relationship between technological progress and the demand for labor does not necessarily hold for craftsmen because, unlike land, the de-trended capital stock can increase independently from the pace of technological progress. If the capital stock at least keeps pace with technological change, the demand for labor in manufacturing will not decrease for a given wage.

Also note that a higher relative price of agricultural goods increases the demand for farm hands, raising the wage in manufacturing units, and lowers the demand for labor by the craftsmen. This is one of the ways that higher agricultural prices causes de-industrialization—it reduces the fraction of the workforce in the manufacturing sector.

8.2.6 Capital Market Equilibrium

The supply of capital, acquired for retirement saving by the three household types, must equal the demand for capital by craftsmen in the next period,

$$\bar{N}_t s_t + N_t^* s_t^* + \tilde{N}_t \tilde{s}_t = N_{t+1}^* k_{t+1}. \tag{8.22}$$

We assume that the population of each type increases at the same rate. The common population growth factor is n. Using the saving equations for each type, the demand for labor expressions, and the labor market equilibrium condition, we can derive the following transition equation for the de-trended capital stock,

$$\widehat{k}_{t+1} = \frac{\beta w_t}{(1+\beta)(1+d)n}\left[\frac{\bar{N}_t}{N_t^*} + \frac{w_t^*}{w_t} + \frac{\tilde{N}_t}{N_t^*}\left(1 - \frac{\alpha \qquad \text{rent}_{t+1}}{(1+\beta)(1-\alpha)w_t D_t R_t}\right)\right]. \tag{8.23}$$

The transition equation is similar to the standard ones derived in simpler one sector overlapping-generations models. Next period's capital stock per craftsman is determined by the saving of this period's young households, whose saving are fractions of their first period income. Growth in the effective labor supply of craftsman (determined by d and n) makes it more difficult to raise the capital intensity at any one shop. These familiar mechanisms are captured by the term outside of the squared-bracket.

The terms inside the squared brackets appear because there are three different household types in the model. The first term accounts for the different population sizes of migrant workers and craftsmen. The second term adjusts for the wage premium received by craftsmen. The final term results from the lower saving rate of farmers relative to the other two household types.

An increase in the relative price of agricultural products increases land rents and reduces the capital stock, other things constant. This is the second type of de-industrialization effect from higher agricultural prices—it reduces saving and lowers the capital stock. From (8.21), lowering the capital stock, also serves to reduce the fraction of the workforce employed in the manufacturing sector.

The solution of the model involves solving simultaneously for $\widehat{k}_{t+1}$ and w_{t+1} using (8.21) and (8.28). The Appendix discusses this solution in more detail.

8.2.7 The Urban/Manufacturing Share and Aggregate Output

We will need to keep track of the urban share and the growth in total output in determining our calibration of the model to the Ottoman economic history. We define the urban share as the craftsmen plus the workers they hire divided by the total population,

$$urbanshare_t \equiv \frac{N_t^* + N_t^* \left(\frac{\overline{\alpha}}{w_t}\widehat{k}_t^{\alpha}\right)^{\frac{1}{1-\overline{\alpha}}}}{N_t^* + \bar{N}_t + \tilde{N}_t}. \tag{8.24}$$

Total output is the sum of the manufacturing goods produced by craftsmen and agricultural goods produced by farmers,

$$N_t^* y_t + \tilde{N}_t p_t O_t = N_t^* D_t \widehat{k}_t^{\alpha} \left(\frac{\overline{\alpha}\widehat{k}_t^{\alpha}}{w_t}\right)^{\frac{\overline{\alpha}}{1-\overline{\alpha}}} + \left(\frac{1-\alpha}{w_t}\right)^{\frac{1-\alpha}{\alpha}} p_t^{1/\alpha} L. \tag{8.25}$$

Notice in (8.25) that technological change only increases output in the manufacturing sector. Why is this the case, given that we assume balanced technological change throughout the economy?

Going back to (8.21), we see that technological progress reduces the demand for labor in the agricultural sector proportionally. Other things constant, farmers need to employ a certain quantity of *effective* labor to profitably produce agricultural goods. A higher level of technology, increases the productivity of a worker and reduces the number of workers needed. Thus, technological change frees labor from agriculture but does not directly increase output.

In the manufacturing sector, technological progress does not necessarily reduce employment because the demand for labor is also determined by the de-trended capital stock which, unlike land, can be increased with sufficient saving. So, while technological progress is balanced throughout the economy, the larger is the manufacturing sector, the stronger is the effect of technological progress on total output. What is "special" about the manufacturing sector is that it relies more heavily on manmade physical capital. This means that technological progress in manufacturing does not just reduce labor cost, but may also increase output. Thus, there is mechanism through which de-industrialization, in the sense of a lower manufacturing share or a lower capital stock, can reduce economic growth.

8.2.8 Calibration

Each period in the model last 20 years and we think of the initial period as 1820. We set the capital share to the commonly chosen value of 1/3. We set the exogenous population growth rate to 1% annually based on the estimate from Issawi (1980). Also from Issawi, we target urban population shares equal to 17% in 1820 and 22% in 1920. We assume that a constant 10% share of the population is comprised of craftsmen. This means that the percent of the population that are migrant workers living and working for the craftsmen in the city begins at 7% and then grows to 12% by 1920. We targeted an initial interest rate at the beginning of the century of 20% to

Table 8.4 Nineteenth century Ottoman Empire

Time period	Wage (manufacturing units)	Interest rate (percent)	Urban share (percent)	Skill premium	Farmer's land rent (manufacturing units)
1820	0.025	20.2	17	1.50	0.64
1840	0.035	15.0	19	1.82	0.92
1860	0.043	14.0	17	1.56	1.21
1880	0.039	12.9	19	2.04	1.16
1900	0.036	12.9	21	2.32	1.11
1920	0.034	12.9	22	2.55	1.07

capture the lack of advancement in financial intermediation and the very high interest rates in the Empire over the century. Given the absence of data, we somewhat arbitrarily target an initial craftsman skill premium of 1.5. The four targets: initial and end of century urbanization, initial interest rate, and initial skill premium, determine $\beta = 0.065, \overline{\alpha} = 0.21,$ and the initial values for the de-trended stocks of land and physical capital.

The exogenous rate of technological progress is set to keep the annual growth in per capita income between 0.5 and 1% over the century, based on the estimates from Pamuk (2006). The exogenous rate of technological change was set to 0.3% annually. The initial value for the international relative price of agriculture goods is one. In 1840 the relative price rises to 1.5 and then to 2.0 in 1860, after which it is kept constant.[8]

Table 8.4 gives the result of the calibrated development of the nineteenth century Ottoman Empire. The associated annual growth rate in per capita income was 0.78%. The reported wage rate is w, i.e. the wage de-trended for technological progress.

Clearly the jump in agricultural prices in the middle of the century slowed the structural transformation. The urban share and the skill premium both declined, before rising again later in the century. We won't actually know the full extent of the slowdown until we compute the counterfactual simulation where we assume there is no rise in the international price of agricultural goods. Land rents rose and then fell as agricultural prices stabilized at their higher value and the land per farm fell. The rise in land rents increased the demand for labor, driving up the wage in manufacturing units. As we shall see below, the higher wage is the main reason for the slow growth in urban employment as a share of the population.

[8]We could have chosen a greater rise in the relative price but did not for the following reasons. As stressed by Quataert (1993), the rural sector, as is typically the case in developing economies, actually engaged in some informal manufacturing production. This means that the relative price of rural production did not rise as much as the relative price of food and primary commodities. We found that, with greater increases in the relative price of rural production, it was difficult to match the 1920 urbanization target without also assuming greater growth than was consistent with Pamuk's findings (growth in per capita of less than one percent). A better way of addressing these points would be to allow the farmers to produce all types of goods, but we leave this for future work.

Table 8.5 Counterfactual Ottoman Empire (constant p)

Tint period	Wage (manufacturing units)	Interest rate (percent)	Urban share (percent)	Skill premium	Land rent (per farmer)
1820	0.025	20.2	17	1.50	0.64
1840	0.024	13.9	27	3.64	0.57
1860	0.023	13.4	30	4.35	0.54
1880	0.021	13.3	33	4.86	0.51
1900	0.02	13.4	35	5.37	0.49
1920	0.0 18	13.5	38	5.92	0.46

8.2.9 Counterfactual Simulation

Now we compute the counterfactual development path, assuming the relative price of agricultural goods remains constant at one throughout the century. The results are given in Table 8.5. The effect on urbanization is dramatic; instead of 22%, the urban share hits 38% in 1920. The annual growth rate in per capita income increases modestly to 0.91% from 0.78% per year. The higher growth rate is due to a rise in capital intensity and a shift of the labor to the modern production, both of which raise the impact of exogenous technological progress. The qualitative predictions of the De-industrialization hypothesis clearly hold.

However, there are distributional effects that make it unclear whether the majority of the population was in fact hurt by the rise in agricultural prices. There is a dramatic rise in the skill premium in the counterfactual simulation, indicating that the craftsmen would have been much better off if the price rise had not occurred. Farmers/landowners would have seen a decline in land rents rather than the rising rents that they actually experienced. In addition, migrant workers would have seen a decline in their real wage in manufacturing units rather than the significant rise we see in Table 8.4. On the other hand, migrant workers would not have confronted the higher price of agricultural goods. To draw clear conclusions, we need a more complete metric for the overall welfare effects.

To compare the welfare effects associated with the growth paths from Tables 8.4 and 8.5, we compute the value function for each household type.

Migrants

$$V_t^M = \Gamma_t + (1+\beta)\ln w_t + \beta \ln R_t - (1-\chi)\left(\ln p_t + \beta \ln p_{t+1}\right)$$

Craftsmen

$$V_t^C = \Gamma_t + (1+\beta)\ln w_t^* + \beta \ln R_t - (1-\chi)\left(\ln p_t + \beta \ln p_{t+1}\right)$$

Farmers

$$V_t^F = \Gamma_t + (1+\beta)\ln w_t + \beta \ln R_t - (1-\chi)\left(\ln p_t + \beta \ln p_{t+1}\right) + (1+\beta)\ln\left(1 + \frac{rent_{t+1}}{w_t D_t R_t}\right).$$

The expression Γ_t is comprised of parameters and exogenous variables that are independent of the endogenous variables along the growth path. The welfare effects are obviously going to depend on the fraction of budget spent on food, $1-\chi$. In a developing economy the food share will be high. We compute the welfare difference between the counterfactual case and the historical simulation for $1-\chi$ set to ¾.

Table 8.6 reports the lifetime welfare effects for each generation of our three household types. There are small differences in the return to capital across the historical and counterfactual simulations, so the small resulting differences in welfare due to changes in the return to capital will be ignored in our discussion.

For the initial generation of migrants and craftsmen, eliminating the price-rise makes them better off. The price increase did not begin until 1840 and so did not affect the wages of the young workers in 1820. However, the higher price of food lowers purchasing power for these households when they are old and lowers their lifetime welfare. The older farmers received more than enough additional land rents in 1840 to cover the higher price of food, so they would be worse off had the price rise been eliminated. Not surprisingly, all generations of craftsmen are made better-off, and all generations of farmers worse-off, when eliminating the relative price increase.

Starting in 1840, eliminating the rise in prices also affects wages. Wages would have been lower if the strong demand for labor by the agricultural sector was eliminated. The main result from Table 8.6 is that, starting in 1840, migrant workers would have been worse off had the international price of food not risen. The rise in their wages was enough to offset the rise in the price of food, even when assuming that 75% of their expenditures were on food. This is because without the rise in p, the de-trended wage in manufacturing units would have fallen by 29% instead of rising by 33%. By century's end, wages were 86% higher with the rise in p than without it. This is more than enough to compensate for the rise in p.

After 1840, only craftsman would have benefited, had the Empire not been exposed to the price rise in international markets. Skilled wages, or monopoly rents, would have been much higher had the relative price not turned so dramatically

Table 8.6 Welfare effects from counterfactually eliminating the rise in p

Time period	Migrants	Craftsmen	Farmer/landowners
1820	0.0073	0.0073	−0.0422
1840	−0.0587	0.6794	−0.119
1860	−0.1191	0.973	−0.1616
1880	−0.1157	0.8088	−0.1616
1900	−0.1066	0.7918	−0.1595

against the goods that craftsmen produced. Thus, while economic growth would have been somewhat higher had the relative price rise not occurred, only a very small fraction of the society would have benefited.

This conclusion ignores the negative effects of an increase in the volatility of relative prices, “terms-of-trade,” due to opening the economy. Blattman et al. (2007) find that opening the economy increased volatility for developing economies in the 19th and early 20th centuries. They also estimate that an increase in terms-of-trade volatility is associated with slower growth in per capita income. Increased volatility can also reduce welfare directly. A more complete analysis of opening the economy should include the potentially important effects of increased volatility in the terms-of-trade, as we discuss in the next section.

8.3 Other Theories of Trade and Growth

In Sect. 8.2 we have seen that international trade may lower growth in developing countries that have a comparative in traditional agricultural production. In this section we discuss some other arguments suggesting that trade may lower growth.

8.3.1 Reduced Incentives for Human Capital Investment

A closely complimentary theory about how trade may lower growth in developing countries is offered by Galor and Mountford (2006, 2008). They argue that higher relative prices in the traditional sector will lower the demand for skilled labor and reduce human capital accumulation. In our formal model, primarily for simplicity, the evolution of human capital is independent of the structural transformation. This is because we have exclusively emphasized how the reliance of families on child labor affects their willingness to educate children.

A second consideration involved in the decision to educate children is the relative wage paid to skilled labor versus unskilled labor. To analyze this situation, skilled and unskilled labor must be treated as distinct inputs in production. In our model, units of human capital are perfect substitutes, known as “efficiency units,”—it simply takes more units of unskilled labor to do what a skilled worker can do. Alternatively, skilled labor could be an important input in the modern sector, while unskilled labor suffices for traditional sector production. If this is true then an increase in the relative price of traditional sector goods lowers the relative demand for skilled labor and lowers the incentive to attend school. This argument is overly strong because there is evidence that schooling increases farmers’ productivity even in traditional settings. However, the relative return to schooling may nevertheless be higher in the modern sector and the argument will go through more generally.

8.3.2 Efficiency Advantages and External Effects from Urbanization

It is often assumed that there is a productivity advantage of producing manufacturing goods in urban locations rather than producing primary commodities in rural locations (e.g. Williamson 2011, p. 49). The ideas underlying this assumption include the possibility that (1) technological change occurs more rapidly in manufacturing than in agriculture, (2) greater manufacturing production generates economies of scale, and (3) urbanization creates economic externalities and spillovers.

It is quite difficult to directly detect a productivity advantage simply by looking at data across sectors. In the U.S. and Europe, the productivity advantage of urban manufacturing certainly did not show up as large differences in hourly labor productivity of agricultural and nonagricultural workers during the nineteenth century. Measures of productivity and wage gaps per hour worked were quite small, less than 25%, even without adjusting for differences in education per worker. As discussed in Chap. 7, some researchers also find small hourly productivity gaps in current developing countries, adjusted for human capital differences.

Of course, sector-wage differences will motivate labor flows that limit the size of the wage gaps. Small gaps in productivity and wages may be consistent with large advantages of urban manufacturing that cause labor to consistently flow toward cities. Thus, the productivity advantages of producing manufacturing goods in an urban setting may not show up as large or rising wage gaps, but rather as increasing labor shares.

In most countries there is a steady increase in the fraction of the workforce in urban manufacturing as economic growth takes place. However, this observation alone also fails to imply the presence of urban manufacturing productivity advantages like the ones mentioned above. If the rural population is increasing, a movement of labor to the city is necessary to offset a declining marginal productivity of rural labor as the labor-land ratio grows. This force for urbanization simply results from the fact that manmade physical capital can be more easily expanded to accommodate a larger workforce than can land.

Others point to the time series correlation between average worker productivity and urbanization as a country grows, but again the causation is far from clear. Henderson (2003) finds “little support for the idea that urbanization per se drives growth.” Rather he argues that “urbanization is a “by-product” of the move out of agriculture and the effective development of a modern manufacturing sector, as economic development proceeds, rather than a growth stimulus.” In other words, there are fundamental changes that cause *both* manufacturing and growth to increase, such as *balanced* technological change across all sectors or increased capital formation, that do not imply any causal feedback effect from urbanization to aggregate growth. In fact, we saw this in Sect. 8.2. Uniquely rapid growth in manufacturing technologies and endogenous external effects from urbanization are not necessary to explain the time series correlation between growth and urbanization. We discuss the link between urbanization and growth further in Chap. 10.

8.3.3 Power of Anti-growth Landowners

There is a political economy argument that links urbanization to growth. The idea is that in the early stages of development large landowners are often politically influential. Their objective is to maximize land rents from traditional agricultural production. Land rents are high when the supply of labor is plentiful and cheap. Landowners are against policies that increase the demand for labor in other sectors and wages throughout the economy. Thus, landowners will oppose policies that promote physical and human capital formation that impact worker productivity more in manufacturing than in traditional agriculture. The greater is the relative size of the urban population, the more interest there is in promoting capital formation and the more likely the interest of the large landowners can be defeated, setting the stage for more pro-growth policies. We provide an example of how the influence of landowners can affect policy in Sect. 8.4.

8.3.4 Volatility and Growth

There is a literature relating short-run volatility to long-run average growth rates. The studies consistently show greater volatility is associated with lower average growth rates (see the review in Cavalcanti et al. 2012). As mentioned in Sect. 8.2, one important source of volatility is the variation in a country's terms of trade. Blattman et al. (2007) provide evidence that terms of trade volatility was associated with lower growth rates in the 1870–1939 period.

Arbatli (2016) conducts a careful study of the region perhaps most affected by the opening of trade in the nineteenth century—Ottoman Turkey. Pamuk and Williamson (2011) conclude that the Ottoman Empire underwent the greatest change in the terms-of-trade of all the states experiencing expanded trade during the century, including countries such as China and India. The trend changes in the terms-of-trade were large but so was the change in volatility about the trend. The coefficient of variation for the terms of trade (a measure of variation relative to the mean) more than doubled. Arbatli concludes that the increase in volatility was a more important detriment to growth than the rise in the average relative price of non-manufacturing goods—the p from Sect. 8.2. He estimates that Ottoman growth could have been 0.7 percentage points faster if volatility was half its actual value from 1800 to 1870.

In addition to terms-of-trade volatility, openness can increase output volatility. A recent study by Haddard et al. (2012) finds that modern day developing economies that are more open experience more volatility in total production. Much of the increased volatility is generated by increased export concentration. They find that openness and export concentration interact to create greater volatility in production. So, export concentration has a larger impact on aggregate volatility the larger is the country's trade share (export plus imports divided by GDP).

8.4 Large Landowners—Growth and Endogenous Fiscal Policy

Now we suppose that there are two rather different private sector household types—workers and large landowners. The workers might farm land, but they do not own the land. The main motivation for this setup is that large landowners are viewed as having an important impact on the political economy of many developing countries, as mentioned in Sect. 8.3. Landowners derive their income from land rents and thus seek to establish and maintain conditions where land rents are high. This motivation comes in conflict with economic progress that raises wages and the cost of labor, so large land owners tend to support policies that stifle economic growth. In this section, we focus on landowner support for fiscal policies that serve their interest.

The dual economy approach also allows us to examine the connection between development and the size of government. Several studies have found a strong negative correlation between the relative size of the agricultural sector and the relative size of government, other things constant (Burgess and Stern 1993; Peltzman 1980; Stotsky and WoldeMarian 1997; Tanzi 1991). In fact, the studies find that the relative size of the agricultural sector is more closely correlated with the relative size of government than are other indicators of development, such as income per capita. One reason for this negative correlation is that the traditional sector generates unrecorded sales and income that are relatively difficult to tax.

The political influence of large landowners is one factor that keeps both the modern sector and the size of government small. There is a growing literature suggesting that land inequality may hamper growth. The survey by Erickson and Vollrath (2004) mentions general mechanisms for the negative effects of land inequality that work through institutions, influence over agricultural policy, credit market development, and support for public schooling. A common feature of the mechanisms is the attempt by politically powerful landowners to maintain a low-cost work force in agriculture by limiting the options of workers outside of agriculture (see Burgress and Stern 1993 for some specific examples from Latin America).

We argue that an additional way that landowners might maintain a low-cost work force is to support high tax rates on labor and capital. If incomes are easier to identify and tax in urban manufacturing, then a high tax-rate environment will favor the agricultural sector. As workers avoid high tax rates by supplying labor to agriculture, the wage rates in agriculture will be driven down to the benefit of landowners. Thus, landowner support for high tax rates will reduce the modern sector, the tax base, and the size of government.

8.4.1 Households

We continue to assume that all households have the same preferences,

$$U_t = \ln c_{1t} + \beta \ln c_{2t+1}. \tag{8.26}$$

Households supply one total unit of labor with no explicit labor/leisure choice. There is no land market and landowners pass their land holdings to their children but derive no explicit utility from doing so.

All working households are now landless and derive their income from supplying labor to both the modern and traditional sectors during the first period of their lives, so they can move across sectors to work without cost. They retire in the second period. Their only source of lifetime income is ω_t, after-tax wage income. For simplicity, we assume the government is completely unable to tax wages earned in the traditional sector. All results go through with a partial ability to tax the traditional sector. After-tax wage income is the sum of after-tax wages earned by workers in each sector, $\omega_t = \pi_t(1-\tau_t)w_t + (1-\pi_t)\tilde{w}_t$, where π is the fraction of work effort supplied to the modern sector.

If both sectors are to operate, workers must be indifferent about where they work. This means that after-tax wages must be equalized across sectors, $(1-\tau_t)w_t = \tilde{w}_t$. Thus, we have a wage gap in *before-tax* wages, resulting from taxation in the modern sector only.

Landowners have the same preferences as workers. In this section we assume the landowners derive *first* period income from the residual income generated by traditional production. This income may be interpreted as a combination of land rents and compensation for the landowner's work time. Land is then passed to the next generation inter vivos at the end of the first period. This timing of the land transfers allows us to bypass the effect of inheritance on the landowners saving rate because it is no longer a source of retirement income.

The landowners lifetime income is then $O_t - \tilde{w}_t f_t$, where f refers to the demand for farm labor. Using the first order condition for the labor demand that maximizes residual income, allows us to write landowner lifetime income as

$$\rho O_t = \frac{\rho \tilde{w}_t f_t}{1-\rho}. \tag{8.27}$$

The demand for labor is given by a new version of (8.8b),

$$f_t = \left[\frac{(1-\rho)}{(1-\tau_t)w_t}\right]^{\frac{1}{\rho}} l_t. \tag{8.28}$$

Note that, if we define the number of landowners as $\tilde{N}_t$, then $\tilde{N}_t f_t = (1-\pi_t)N_t$, where N continues to denote the total number of young working households. So, the fraction of the work force in the modern sector is

$$\pi_t = 1 - \frac{L}{N_t}\left[\frac{1-\rho}{(1-\tau_t)\mathrm{w}_t}\right]^{\frac{1}{\rho}}.$$

8.4.2 Open Economy

For simplicity, as in the later portions of Chaps. 2, 3, and 5, we assume that the economy is open to international capital flows. This assumption forces the domestic rate of return to capital to equal the exogenous world rental rate, $r_t = r*$. The modern sector is as described in Sect. 8.1. The capital-labor ratio in the modern sector is therefore also fixed (see 8.2b) at a value we call k^*, causing the before-tax wage in the modern sector to be fixed at w^*(see 8.2a). This means the welfare of all households is completely driven by the wage tax. In addition, the total capital stock of the country will vary with the size of the modern sector as defined by π_t because $K_t = k^* \pi_t N_t$. Policies that reduce labor in the modern sector will cause the economy to lose physical capital or de-industrialize.[9]

8.4.3 Government Policy

We take the same "reduced-form" approach to the formation of policy that was introduced in Chap. 3. We model the government as any other economic agent—by specifying its preferences, constraints, and objectives. There is no deep model of the politics that determine how the government is chosen and how its policies are influenced by voters and interest groups. Instead we take as given the politics of a country which determine the "reduced-form" preference parameters that dictate the government's concern with the welfare of the general population, of different household-types, and of the households that make up the government itself.

Government officials have preferences defined over their own consumption and the welfare of the two private-sector household types, based on their political influence.

The government officials retire in the second period just as the private agents. Their first period wages are financed by taxes on the wages of the private sector workers. The single period government budget constraint is $w_t^g N_t^g = \tau_t w_t \pi_t N_t$, where all government consumption is in the form of wages paid to the officials and where N_t^g is the number of government officials. The number of government officials is an exogenous fraction of the total population.

The preferences that determine government policy are given by the function, $\phi^g V_t^g + \phi V_t + \widetilde{\phi} \widetilde{V}_t$, where ϕ^g, ϕ and $\widetilde{\phi}$ are constant preference parameters that are

[9]Taxing the return to capital, in addition, to wages would not alter the results much. In an open economy, the after-tax return to capital must remain equal to the after-tax world interest rate. Thus, country specific taxes cannot alter the after-tax return. However, higher taxes on capital in a given country will reduce that country's capital-labor ratio. Thus, taxing capital in an open economy will be entirely shifted to labor by lowering before-tax wages. The primary difference between an income tax and a wage tax is that the economy reduces its capital-labor ratio as well as its total capital stock. See *Problem* 12.

determined by the political power of the three agents and the V functions are the indirect utility functions of each type. We think of the government's preference parameters as functions of exogenous political institutions and the de facto political power of the private sector households. Countries with less democratic institutions, and fewer "constraints on the executive," will tend to have governments that place less weight on the welfare of the private sectors households as a whole (low values for both ϕ and $\tilde{\phi}$), or perhaps that give disproportionate influence to wealthy landowners (a high value for $\tilde{\phi}$ and a low value for ϕ).

The indirect utility functions of the households can be written out as

$$V_t = E + (1+\beta)\ln(1-\tau_t) \tag{8.29a}$$

$$\tilde{V}_t = \tilde{E} - (1+\beta)\frac{1-\rho}{\rho}\ln(1-\tau_t) \tag{8.29b}$$

$$V_t^g = E^g + (1+\beta)(\ln\tau_t + \ln\pi_t), \tag{8.29c}$$

where the upper-case E-expressions on the right-hand-side of each equation contain exogenous constants that will not affect the policy choice. The tax rate lowers the welfare of workers by lowering the after-tax wage in the modern sector and the before-tax wage in the traditional sector. Landowners prefer a high wage tax because it lowers the cost of labor and increases total land rents. The government officials also benefit from a high tax, although they must consider that a higher tax rate lowers the tax base—i.e. the total wage bill in the modern sector.

Using (8.29a, 8.29b, 8.29c) and the equilibrium condition for the labor share, determined from (8.28), the first order condition for the optimal tax rate is

$$\frac{\phi^g}{\tau_t} = \frac{1}{(1-\tau_t)}\left\{\phi^g\left(\frac{1-\pi_t}{\rho\pi_t}\right) + \phi - \frac{1-\rho}{\rho}\tilde{\phi}\right\}, \tag{8.30}$$

which is sketched in Fig. 8.1.

The left-hand-side of (8.30) is the decreasing marginal benefit of taxation that stems from the marginal utility of consumption by government officials whose salaries are financed by the tax revenue. The right-hand-side of (8.30) is the increasing marginal cost of taxation, comprised of three distinct terms. The first term captures the effect of raising the tax rate on the tax base. A higher tax rate shrinks the taxable wage bill in the modern sector as workers move to the traditional sector to avoid taxation. If there was no weight placed on the welfare of private sector households ($\phi = \overline{\phi} = 0$), the government would maximize the tax revenue collected by equating the left-hand-side to the first term on the right-hand side

The second term on the right-hand-side captures the marginal cost of taxation to working households resulting from a reduction in their after-tax wage. The third term, reduces the marginal cost of taxation, because it represents the gain to landowners from the fall in traditional sector wages when the tax rate is increased. If the sum of these last two terms is positive, the optimal tax will be less than the tax

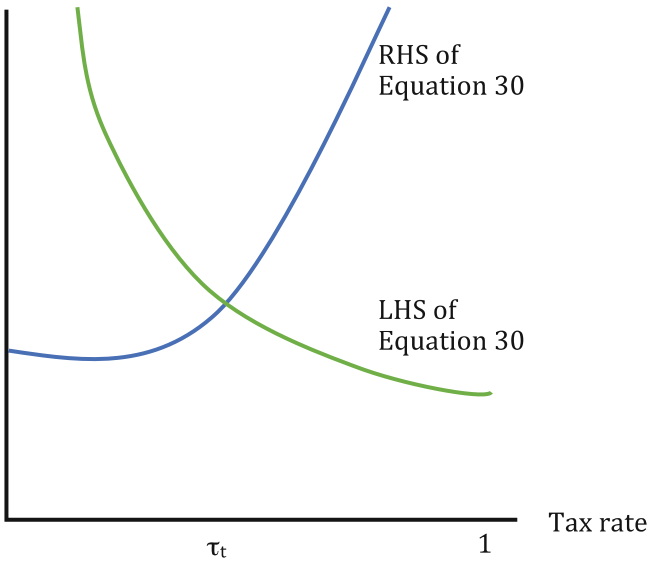


Fig. 8.1 Optimal tax rate

rate that maximize tax revenue because of the net welfare loss that taxes inflict on the private sector. However, with sufficiently powerful landowners the net welfare effect of taxation on the private sector could be positive. In this case, the tax rate would exceed the revenue maximizing level.

There are two important general points demonstrated by (8.30). First, the greater the political influence of landowners (higher $\widetilde{\phi}$), the higher is the tax rate. A higher tax rate lowers wages, the size of the modern sector, and the economy's capital stock. Powerful landowners prevent industrialization of the economy at the expense of working households.[10]

Second, exogenous factors that cause the size of the modern sector to grow, i.e. that cause π_t to increase for a given tax rate, lower the marginal cost of taxation and cause the optimal tax rate to increase. The intuition is that the *marginal* loss in the tax base, as the tax rate rises, is smaller and less valuable, the larger is the *total* tax base. Thus, tax rates will tend to increase, other things constant, as economies grow and modernize. This result helps explain Wagner's Law, the observation that the relative size of government increases with development.[11] Some evidence for Wagner's Law is provided below in Table 8.7.

[10]Galor et al. (2009) provide a theory and supporting evidence that larger landowners have acted to slow the accumulation of human capital for similar reasons.

[11]Our analysis ignores the growth in the size of government due to the growth in social transfers. This reason for the growth in government tends to occur in later stages of development as countries become more democratic. See Lindert (2004) for a thorough discussion of the connection between democracy and government size.

Table 8.7 Government in developed and developing countries

	Government purchases (% of GDP)	Capital income tax rate (maximum statutory rate)	Personal income tax rate (maximum statutory rate)
Developed countries (1990s)	18.9	29.6	42.8
Developing countries (1990s)	14.2	26.7	34.7
Developing countries (1870)	4.6	n.a.	n.a

Sources: Government purchase share for 1990 from Jha (2007. Table 9). Government purchase share for 18 70 from Tanzi and Schuknecht (2000. Table II.l). Tax rates for 1996–2001 ate from Gordon and Li (2005)

The rise in the relative size of the government over the course of development is associated with constant or rising economic growth rates. We have seen that taxation can reduce private capital accumulation. What explains this apparent paradox? The answer is that the government uses some fraction of rising tax revenue to invest in public education, public health, and infrastructure, as we saw in Chap. 3. As the structural transformation generates a relatively larger government, there need not be a drag on growth if the government uses a sufficiently high fraction of the tax revnue on investment. Mourmouras and Rangazas (2009b) discussed these points in detail.

8.5 Conclusion

In this chapter we see that the structural transformation expands hours worked per worker because workers are constrained to work significantly fewer hours per year in the traditional agricultural sector. Agricultural workers are willing to accept this constraint because they are compensated by the rents they receive from the land they inherit from their parents. The structural transformation has ambiguous effects on physical capital per worker. Migration out of the traditional sector has two opposing effects on capital intensity; saving rates rise but the number of workers that must be supplied with capital in the modern sector also rises.

In the U.S. we find the structural transformation in the nineteenth century did not have a significant impact on capital intensity but did significantly expand hours worked per worker. This resulted in a much larger increase in output per worker than in output per hour worked.

We also examined the extent to which international trade “de-industrialized” the Ottomon Empire in the nineteenth century. We find that the sharp rise in the relative price of agricultural products did significantly slow the migration of labor into the urban manufacturing sector. There was also a small negative impact on labor productivity per hour worked. As in the case of the U.S., the pace of structural transformation is not closely related to the growth in capital intensity. De-industrialization did not cause significant changes in capital intensity. In addition, we showed that it is possible for the vast majority of the population to have benefited from the rise in the relative price of agricultural goods. Incorporation of the effect of greater volatility in the term-of-trade is needed to provide a more conclusive assessment.

While the results from this chapter suggest a weak connection between private capital accumulation and the structural transformation, there is an important connection to public capital accumulation. Factors that reduce the size of the traditional sector increase the government ability to raise tax revenue. The rise in tax revenue increases the funding for expanding public investment projects in countries with pro-growth governments.

8.6 Exercises

Questions

1. Explain how landed and landless households might differ in terms of (a) preferences, (b) constraints, and (c) behavior.
2. How does the structural transformation affect physical capital accumulation when many smaller farmers own land? Compare the theory of capital formation in Sect. 8.1 to that in Chap. 4 where land markets are assumed to be perfectly competitive.
3. Explain why the structural transformation affects labor productivity growth differently depending on whether it is measured per hour or per worker. How does U.S. economic growth indicate that this distinction is important?
4. Describe the three types of households in Sect. 8.2. What is the motivation for introducing these particular household types?
5. In the model in Sect. 8.2, explain why it is possible to interpret the income of the craftsmen as either the "skilled-worker's wage" or as "monopoly profit." Under what condition is the craftsman's income greater than the wage of the migrant worker?
6. Explain two ways in which international trade might "de-industrialize" an economy. Based on the analysis in Sect. 8.2, did international trade "de-industrialize" the Ottoman Empire in the nineteenth century?
7. *Thinking about* Table 8.4. (a) Use (8.21) to explain why wages in manufacturing units rise and then fall (b) Use (8.17a) to explain why the decline in interest rates implies the de-trended capital stock per craftsman must rise (c) Use (8.17c) to explain the time path of the skill premium (d) Why does a farmer's land rent rise and then fall?
8. Based on the analysis in Sect. 8.2, how did international trade alter growth and economic welfare in the Ottoman Empire over the nineteenth century?
9. Discuss alternative theories of how international trade affects economic growth. Do these theories compliment or compete with the theory in this chapter?

10. Is encouraging developing economies to open their economies a good idea? Base your answer on the analysis from this chapter and Chap. 5.
11. How are each of the three households in 8.3 affected by wage taxation?
12. Explain how powerful landowners can slow the structural transformation through their influence of tax policy.
13. Give two explanations for Wagner's Law.
14. Use Fig. 8.1 to explain your answers to questions 12 and 13.
15. How does an increase the political power of the government officials affect tax policy? Careful—this is a bit harder than it appears.
16. Use the models of this Chapter *G9* and *G12*.

Problems

1. Derive and interpret the saving function of a young household that eventually inherits the family farm as in (8.8a).
2. Derive the indirect utility functions of landed and landless households and use them to derive (8.9). Show that Ω_{t+1} is decreasing in $\tilde{z}$.
3. Show that (8.9) can be written as a function of π_t. So given k_t and N_t, one can use (8.9) to solve for the split of the population across sectors.
4. Derive the transition equation for physical capital intensity in 8.1. Carefully explain how one could simulate the transition path of the economy during the structural transformation.
5. Derive the labor-market equilibrium condition from Sect. 8.2. How does technological change affect the demand for labor, other things constant?
6. Derive the expression for total output from Sect. 8.2. Why does the impact of technological change on output depend on the fraction of the work force in the modern sector?
7. In the Appendix we claim that an increase in the supply of unskilled labor ($\bar{N}_{t+1}+\tilde{N}_{t+1}$) lowers the present value of land rents relative to wages and raises saving. To see this first show that $rent_{t+1}/w_t D_t R_t = \frac{\alpha}{1-\alpha}\frac{(1+d)f_{t+1}}{w_t}\frac{w_{t+1}}{R_t}$. Next, use (8.17a, 8.17b, 8.17c) and (8.20) to show $\dfrac{w_{t+1}}{R_{t+1}} = \dfrac{w_{t+1}}{r_{t+1}} = \dfrac{\overline{\alpha}}{\alpha}\dfrac{N^*_{t+1}}{\left(\tilde{N}_{t+1}+\bar{N}_{t+1}\right)-\tilde{N}_t f_{t+1}}$.
8. Use the information in Tables 8.4 and 8.5 to compute the value of χ that would make each generation of migrant workers indifferent about the rise in the relative price of the agricultural good. Why is it valid to use the *same* values computed in Tables 8.4 and 8.5, when considering *different* values of χ?
9. Derive the indirect utility function for each household in 8.4.
10. Show that $\frac{d\pi_t}{d\tau_t} = \frac{-1}{\rho}\frac{1-\pi_t}{1-\tau_t}$. So the *marginal* decline in the modern sector tax base is smaller the larger is the tax base.
11. Derive the equation determining the optimal wage tax rate in Sect. 8.4.
12. Suppose that instead of a wage tax, the government imposes an income tax that taxes both wages and the return to capital, r_t. Redo the analysis in Sect. 8.4 under the income tax.

Appendix

Here are a few notes about solving for the dynamic path of the model from Sect. 8.2. The transition equation given by (8.23) can be written out in terms of wages and capital intensity,

$$\widehat{k}_{t+1} = \frac{\beta w_t}{(1+\beta)(1+d)n}\left[\frac{\dfrac{\dfrac{\bar{N}_t+\tilde{N}_t}{N_t^*}+1-\overline{\alpha}-\alpha}{\overline{\alpha}}\left(\dfrac{\overline{\alpha}\widehat{k}_t^{\alpha}}{w_t}\right)^{\frac{1}{1-\overline{\alpha}}}}{1+\dfrac{1}{1+\beta}\dfrac{\overline{\alpha}}{1-\alpha}}\frac{\left(\frac{(1-\alpha)p_{t+1}}{w_{t+1}}\right)^{1/\alpha}(\mathrm{L}/D_{t+1})}{\bar{N}_{t+1}+\tilde{N}_{t+1} - \left(\frac{(1-\alpha)p_{t+1}}{w_{t+1}}\right)^{1/\alpha}(\mathrm{L}/D_{t+1})}\right]. \tag{8.23}$$

Given an initial value for $\widehat{k}_t$, one can use (8.21) to solve for w_t. Next, given the values for $\widehat{k}_t$ and w_t, use (8.21), updated one period, and (8.23) to solve simultaneously for $\widehat{k}_{t+1}$ and w_{t+1}.

It is important to note from (8.21) and (8.23), that the split of the population across migrant workers and farmers does not affect the equilibrium determination of the capital stock or wages ($\bar{N}_t$ and $\tilde{N}_t$ only appear together as a sum). In the numerator of (8.23), the farmers and migrants can be aggregated because they receive the same wage. In the denominator, the supply of farmers and migrants minus the farm demand for labor determines the equilibrium employment of unskilled labor in the urban sector (see 8.21). More unskilled labor in the urban sector lowers the capital to unskilled labor ratio and the wage-interest rate ratio (see 8.17a, 8.17b, 8.17c). A lower wage-interest rate ratio lowers the present value of land rents relative to wages, thereby increasing the saving of farmers and capital formation (see Problem 7 for more details).

The fact that $\bar{N}_t$ and $\tilde{N}_t$ only appear together as a sum is important because ownership claims on land in the early stages of development are unclear and it is difficult to determine what fraction of the population were migrant workers and what fraction earned some land rent. In addition, note that the capital stock and factor prices can be determined independently of the composition of output between manufacturing and agricultural goods (i.e. 8.21, 8.22, and 8.23) are independent of χ).

References

Acemoglu D, Robinson J (2012) Why nations fail. The origins of power, prosperity, and poverty. Crown Publishers, New York.

Alston L, Hatton T (1991) The earnings gap between agricultural and manufacturing laborers, 1925–1941. Journal of economic history 51(1):83–99.

Arbatli, C (2016) Trade and income growth in the Ottoman Empire: Assessing the role of volatility and trend growth in the terms of trade. Eurasian Economic Review, 6, 173–194.
Banerjee A, Duflo E (2011) Poor economics. Public Affairs, New York.
Burgess R, Stern N (1993) Taxation and development. Journal of economic literature 31 (2):762–830.
Blattman C, Hwang J, Williamson J (2007) The impact of terms of trade on economic development in the periphery, 1870–1939: Volatility and secular change. Journal of development economics 82(1):156–179.
Carter S, Ransom R, Sutch R (2003) Family matters: The life-cycle transition and the unparalleled fertility decline in antebellum America. In: Guinnane, T Sundstrom W, Whately W (eds) History matters: Essays on economic growth, technology, and demographic change. Stanford University Press, Stanford, California.
Caselli, F, Coleman W (2001) The US structural transformation and regional convergence: A reinterpretation. Journal of political economy 109(3):584–616.
Cavalcanti T, Mohaddes K, Raissi M (2012) Commodity price volatility and the sources of growth. IMF working paper 12/12. International Monetary Fund, Washington DC.
Clark G (2007) A farewell to alms. Princeton University Press, Princeton, New Jersey.
Collier P, Radwan S, Wangwe S (1986) Labour and poverty in rural Tanzania: Ujameen and rural development in the United Republic of Tanzania. Oxford University Press, New York.
Cunningham H (1990) The employment and unemployment of children in England c. 1680–1851. Past and present 126:115–150.
David P (1967) The growth of real product in the United States before 1840: New evidence, controlled conjectures. Journal of economic history 27(2):151–197.
Deninger K, Squire L (1997) Explaining agricultural and agrarian policies in developing countries. Journal of economic literature 35(4):1958–2005.
de Vries J, van de Woude A (1999) The first modern economy. Cambridge University Press, Cambridge, UK.
Eaton J (1987) A dynamic specific factors model of international trade. Review of Economic Studies 54:325–338.
Erickson, L, Vollrath D (2004) Dimensions of land inequality and economic development. IMF Working Paper 158. International Monetary Fund, Washington DC.
Galor O, Mountford A (2006) Trade and the great divergence: The family connection. American economic review 96(2):299–303.
Galor O, Mountford A (2008) Trading population for productivity: Theory and evidence. Review of economic studies 75(4):1143–1179.
Galor O, Moav O, Vollrath D (2009) Inequality in landownership, the emergence of human capital promoting institutions, and the great divergence. Review of economic studies 76(1):143–179.
Goldin C, Sokoloff K (1984) The relative productivity hypothesis of industrialization: The American case 1820–1850. Quarterly journal of economics 99(3):461–490.
Gollin D (2002) Getting income shares right. Journal of political economy 110(2):458–474.
Gollin D (2014) The Lewis model: A sixty year retrospective. Journal of economic perspectives 28 (3):71–88.
Gollin D, Lagakos D, Waugh M (2014) The agricultural productivity gap. Quarterly journal of economics 129(2):939–993.
Gordon R, Li W (2005) Tax structure in developing countries: Many puzzles and a possible explanation. NBER working paper 11267. National Bureau of Economic Research, Cambridge, Massachusetts.
Haddard M, Lim J, Pancaro C, Saborowski C (2012) Trade openness and growth volatility. ECB working paper 1491. European Central Bank, Frankfurt.
Hatton T, Williamson J (1992) What explains wages between the farm and city? Exploring the Todaro model with American evidence, 1890–1941. Economic development and cultural change 40:267–294

Henderson V (2003) The urbanization process and economic growth: The so-what question. Journal of economic growth 8(1):47–71.
Horrell S, Humphries J (1995) The exploitation of little children: Child labor and the family in the industrial revolution. Explorations in economic history 32(4):485-516.
Issawi C (1980) The economic history of Turkey 1800–1914. Chicago University Press, Chicago.
Jha R (2007) Fiscal policy in developing countries: A synoptic view. Australia South Asia Research Centre working paper 2007/01.
Jones R (1971) A three factor model in theory, trade and history. In Bhagwati J et al. (eds) Trade, balance of payments, and growth. North Holland, Amsterdam.
Kendrick J (1961) Productivity trends in the United States. Princeton University Press, Princeton, New Jersey.
La Porta R, Shleifer A (2014) Informality and development. Journal of Economic Perspectives 28(3):109–126.
Lewis A (1954) Development with unlimited supplies of labor. The Manchester School 22:92–139.
Libecap G (2007) Property rights and federal land policy. In: Fishback P (ed) Government and the American economy. University of Chicago Press, Chicago.
Lindert P (2004) Growing public. Cambridge University Press, Cambridge, UK.
Margo R (2000) Wages and labor markets in the United States 1820–1860. University of Chicago Press, Chicago.
Mourmouras A, Rangazas P (2009a) Reconciling Kuznets and Habbakuk in a unified growth model. Journal of economic growth 14(2):149–181.
Mourmouras A, Rangazas P (2009b) Fiscal policy and economic development. Macroeconomic dynamics 13(4):450–476.
Mourmouras A, Rangazas P (2014) Deindustrialization and growth in the Ottoman empire. Working paper, IUPUI, Indianapolis.
Pamuk S (2006) Estimating economic growth in the Middle East since 1820. Journal of economic history 66(1):809–828.
Pamuk S, Williamson J (2011) Ottoman de-industrialization, 1800–1913: Assessing the magnitude, impact, and response. Economic history review 64:159–184.
Peltzman S (1980) The growth of government. Journal of law and economics 23(2):209–287.
Samuelson P (1971) Ohlin was right. Swedish Journal of Economics 73:365–384.
Stotsky J, WoldeMarian A (1997) Tax effects in sub-Saharan Africa. IMF working paper 97/107. International Monetary Fund, Washington DC.
Tanzi V (1991) Structural factors and tax revenue in developing countries: A decade of evidence. International Monetary Fund, Washington DC.
Tanzi V, Schuknecht L (2000) Public spending in the 20th century. Cambridge University Press, New York.
Quataert D (1993) Ottoman manufacturing in the age of the industrial revolution. Cambridge University Press, Cambridge, UK.
Vollrath D (2014) The efficiency of human capital allocations in developing countries. University of Houston Working Paper 201307956.
Williamson J (2011) Trade and poverty: When the third world fell behind. MIT Press, Cambridge, Massachusetts.

A Complete Dual Economy

9

In this chapter we combine the main features from previous chapters into a complete dual economy model. The model includes physical and human capital, fertility, wage gaps, and technological change. We then examine the ability of the model to replicate key features of the economic growth observed in real world economies.

In order to include all the features together in one model, we introduce one simplification. In several of the applications of the two sector model from Chaps. 6, 7, and 8, we interpreted the traditional sector as specializing in agricultural production and the modern sector as specializing in manufacturing production. This interpretation seems quite natural and useful for some purposes, but in this chapter we argue for use of the one-good interpretation of the dual economy. We continue to assume that sectors are distinct because the traditional sector's production is dependent on land and natural resources, while the modern sector's production is capital intensive. In addition, traditional production is conducted by families in a localized and specialized manner, while modern sector production is carried out in factories using general methods.

The family, and not the firm, was the predominant center for production in the United States during the nineteenth century (Ruggles 2001; Carter et al. 2003). The family was a multigenerational producer with assets and management provided by older generations and labor provided by younger generations. Goods were produced not only for home consumption but to sell and trade in the market as well. Moreover, family production was not limited to agricultural products. Manufacturing goods such as leather products, flour, furniture, tools and services such as retail sales were also provided informally.[1]

The quantitative importance of the multigenerational family producer was striking. Ruggles (2001) provides data from the middle of the nineteenth century

[1]Across today's developing countries, high percentages of the poor operate small businesses: 50% for urban poor and 25–98% for rural poor (Banerjee and Duflo 2011, p. 135). La Porta and Schleifer (2014) estimate the informal sector accounts for 30–40% of economic activity in poor countries and an even higher share of employment.

S. Das et al., *Economic Growth and Development*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-319-89755-4_9

United States, showing that 80% of the elderly lived with their adult children. He further demonstrates that the property was owned by the elderly and that the children remained at home to provide the labor needed for production. Family-based production was so prevalent that only about 45% of prime-aged white males were employed in wage and salary jobs.

Land inheritance was certainly an important consideration binding young workers to family production but it was not the only consideration. Young workers also inherited specific skills, local knowledge of productive factors, and business relations based on trust and familiarity with the family. In short, children inherited not only land, but an entire family production technology that was passed down from working with their parents. The following quote from Ruggles (2005, p. 11), again regarding the United States in the nineteenth century, attests to historical relevance of this mechanism, even outside of agriculture.

> In the nineteenth century, the bulk of men in the high-status occupations were proprietors of one sort or another. Many of these people inherited their businesses from their fathers. To a lesser extent, that was true in mid-status jobs as well; among the common titles in that category were bakers, brickmasons, cabinet-makers, carpenters, and shoemakers, who typically had their own shops in that period. Many craftsmen inherited their occupations from their fathers, and a son who lived with his parents was no doubt more likely to inherit. Sales clerks had especially high rates of co-residence in the nineteenth century; many of them probably worked in their fathers' stores with the expectation of eventual inheritance.

Similar views on the nature of traditional production are common. Quataert (1993, p. 2), in his detailed study of rural manufacturing in the nineteenth century Ottoman Empire, makes a strong case for the idea that traditional production generally involved a mix of production activities.

> We now know from studies of the American, British, German, Chinese, and other economies that manufacturing often is an integral, variously important, part of agrarian life. Agrarian economies commonly were mixed and rural families engaged in both agriculture and manufacturing, both for subsistence and commercial purposes.

Ranis and Fei (1961, p. 534), among the original founders of the dual economy approach, used the two-good interpretation in their work. However, they recognized it was not entirely accurate.

> We wish to underscore the absence of any necessary one-to-one relationship between the subsistence sector and agriculture, or between the capitalistic sector and industry in most less-developed economies. The existence of substantial islands of commercialized production in the primary sector and of sizable subsistence enclaves in the small-scale and service industries does not, however, bar Lewis, or us, from using this short-hand terminology.

La Porta and Schleifer (2014) find evidence of an informal sector in the production of *most* goods. They argue that the informal sector in developing countries today is best interpreted as the traditional sector in dual economy models. They document large wage gaps between the informal sector and the more capital-intensive formal sector. An expansion in the formal sector, due to physical and human capital accumulation, gradually pulls labor out of the informal sector over time. This is the same mechanism we focus on in this chapter.

With the broader interpretation of the dual economy, the assumption that land markets do not exist is no longer as critical to the analysis because specific skills, along with local knowledge and business relationships, are sufficient to bind children to the traditional technology. This makes it easier to understand why informal agreements are made to include intergenerational transfers of land and other assets, *even in the presence of land markets*. Land is more valuable if kept within the family because its productivity is tied to the other features of the specific traditional technology that children learn while working for their parents.

We ignore land and use the one-good model to study the sources and patterns of long-run growth, as well as the connection of growth with other features of the economy. Section 9.1 presents the model. Section 9.2 calibrates the model and them simulates long-run transitional growth paths. Section 9.3 compares the general features of the growth simulations to historical and more recent data from developing countries. Section 9.4 presents a case study of the South Korean Growth Miracle. Section 9.5 discusses ways that the modeling of human capital can be expanded to explain more of a developing economy's transitional growth. Section 9.6 reports the empirical evidence regarding convergence of poor to rich countries. Section 9.7 adds to our ongoing discussion of the interaction between politics and the dual economy. Section 9.8 offers an introduction to *three* sector models of the structural transformation needed to study issues associated with the later development of rich countries. Section 9.9 provides a conclusion.

9.1 The Dual Economy

The model in this chapter shares elements from all previous chapters. As in Chaps. 4, 5, and 7 households choose how many children to have and how much time the children spend in school rather than work. As in Chap. 6, there are two sectors with two different technologies. As in Chaps. 7 and 8, there are features that distinguish the traditional and modern sectors that go beyond the technologies used. First, by working for the family in the traditional sector, assets are inherited (the family technology or specific human capital) that allows one to manage family production in the second or "retirement" period of adult life. Second, when young, working in the traditional sector constrains work opportunities. Specifically, there are limits to how many hours one can work over the course of the year because of physical demands of the work, seasonal constraints of the weather, or because of the lack of modern features of the technology used such as electricity. As in Chaps. 2, 6, and 8, we include physical capital accumulation and, in particular, how it is influenced by the structural transformation.

9.1.1 Production Technologies and Factor Prices

Producers in the modern sector are neoclassical firms that use the standard Cobb-Douglas production function

$$Y_t = AK_t^{\alpha}(H_t)^{1-\alpha}, \tag{9.1}$$

where Y denotes output, K is the physical capital stock, H is the effective labor employed in the modern sector and α is the capital share parameter.

The firms operate in perfectly competitive markets for goods and inputs. The standard profit-maximizing factor-price equations for the rental rates on human capital and physical capital are

$$w_t = (1-\alpha)Ak_t^{\alpha} \tag{9.2a}$$

$$r_t = \alpha Ak_t^{\alpha-1}, \tag{9.2b}$$

where $k = K/H$.

Traditional output is produced by a family business that we will continue to refer to as a “farm,” to make a clear contrast between informal production and formal production by firms. The traditional technology used on the farm is

$$O_t = \tilde{A}\left(\tilde{D}_t f_t\right)^{1-\alpha}, \tag{9.3}$$

where O output, f is human capital employed per farm in the traditional sector, and $\tilde{D}_t$ is the index of exogenous labor-augmenting technology in the traditional sector. Note, for simplicity only, we ignore land. As before, we assume there is a market for farm labor. Households can hire labor beyond that provided by the family.

9.1.2 Household Behavior

Preferences are the same for all households and are represented by the utility function used in past chapters, $U_t = \ln c_{1t} + \beta \ln c_{2t+1} + \psi \ln (n_{t+1}h_{t+1})$, and the human capital production function is once again $h_{t+1} = e_t^{\theta}$.

The single period budget constraints for the modern sector households are

$$c_{1t} + s_t + n_{t+1}\eta w_t D_t h_t = w_t D_t h_t + n_{t+1} w_t D_t \gamma \bar{h}(T - e_t) \tag{9.4a}$$

and

$$c_{2t+1} = R_t s_t, \tag{9.4b}$$

where, as before, $R_t \equiv 1 + r_t - \delta$. As in previous chapters we assume that the rate of depreciation is one, for simplicity. The modern household's demand for goods, assets, children, and schooling are

$$c_{1t} = \left[\frac{1}{1+\beta+\psi}\right] w_t D_t h_t \tag{9.5a}$$

$$c_{2t+1} = \left[\frac{\beta}{1+\beta+\psi}\right] R_t w_t D_t h_t \tag{9.5b}$$

$$s_t = \frac{\beta}{1+\beta+\psi} w_t D_t h_t \tag{9.5c}$$

$$n_{t+1} = \frac{\psi}{(1+\beta+\psi)\left(\eta - \gamma(T-e_t)\left(\bar{e}/e_{t-1}\right)^{\theta}\right)} \tag{9.5d}$$

$$e_t = \max\left[\frac{\theta\left(\eta\left(e_{t-1}/\bar{e}\right)^{\theta} - \gamma T\right)}{\gamma(1-\theta)}, \bar{e}\right]. \tag{9.5e}$$

Note that modern sector households are free to work in either the modern or traditional sectors. They can spend part of the year working in the modern sector and part of the year in the traditional sector, and thus they have much more mobility than we assume for the worker that lives in the traditional sector.

There are two fundamental differences between the budget constraints of traditional and modern sector households. First, instead of a full unit of labor, workers who live in the traditional sector must remain on the farm all year and are constrained to work $\tilde{z} < 1$ units of labor time. During down-time in production, the workers stay in the traditional sector to maintain and protect the farm. Second, traditional sector households have residual income from operating the farm in the second period of life, $O_{t+1} - \tilde{w}_{t+1}\tilde{D}_{t+1}f_{t+1}$. The farm technology is jointly operated by all siblings that remain in the traditional sector. The number of siblings from generation-t that operate the farm is some fraction, $\varphi_t \leq 1$, of all siblings ($\tilde{n}_t$). Each of the siblings then passes down knowledge of how to operate a farm to the children who remain in the traditional sector. The children that remain in the traditional sector jointly manage the farm and share in the residual income, as in the previous generation.

The two single-period budget constraints for the traditional households are

$$\tilde{c}_{1t} + \tilde{s}_t + \tilde{n}_{t+1}\eta\tilde{w}_t\tilde{D}_t\tilde{z}\tilde{h}_t = \tilde{w}_t\tilde{D}_t\tilde{z}\tilde{h}_t + \tilde{n}_{t+1}\tilde{w}_t\tilde{D}_t\gamma\tilde{z}\bar{h}\left(T - \tilde{e}_t\right) \tag{9.6a}$$

and

$$\tilde{c}_{2t+1} = R_t\tilde{s}_t + \left(O_{t+1} - \tilde{w}_{t+1}\tilde{D}_{t+1}f_{t+1}\right)/\varphi_t\tilde{n}_t. \tag{9.6b}$$

Traditional household behavior is given by

$$\tilde{c}_{1t} = \left[\frac{1}{1+\beta+\psi}\right]\tilde{w}_t\tilde{D}_t\tilde{z}\tilde{h}_t\left[1+\frac{O_{t+1}-\tilde{w}_{t+1}\tilde{D}_{t+1}f_{t+1}}{R_t\tilde{w}_t\tilde{D}_t\tilde{z}\tilde{h}_t\varphi_t\tilde{n}_t}\right] \tag{9.7a}$$

$$\tilde{c}_{2t+1} = \left[\frac{\beta}{1+\beta+\psi}\right]R_t\tilde{w}_t\tilde{D}_t\tilde{z}\tilde{h}_t\left[1+\frac{O_{t+1}-\tilde{w}_{t+1}\tilde{D}_{t+1}f_{t+1}}{R_t\tilde{w}_t\tilde{D}_t\tilde{z}\tilde{h}_t\varphi_t\tilde{n}_t}\right] \tag{9.7b}$$

$$s_t = \frac{\tilde{w}_t\tilde{D}_t\tilde{z}\tilde{h}_t}{1+\beta+\psi}\left[\beta-(1+\psi)\left(\frac{O_{t+1}-\tilde{w}_{t+1}\tilde{D}_{t+1}f_{t+1}}{R_t\tilde{w}_t\tilde{D}_t\tilde{z}\tilde{h}_t\varphi_t\tilde{n}_t}\right)\right] \tag{9.7c}$$

$$\tilde{n}_{t+1} = \frac{\psi}{(1+\beta+\psi)\left(\eta-\gamma(T-\tilde{e}_t)\left(\bar{e}/\tilde{e}_{t-1}\right)^{\theta}\right)}\left[1+\frac{O_{t+1}-\tilde{w}_{t+1}\tilde{D}_{t+1}f_{t+1}}{R_t\tilde{w}_t\tilde{D}_t\tilde{z}\tilde{h}_t\varphi_t\tilde{n}_t}\right] \tag{9.7d}$$

$$\tilde{e}_t = \max\left[\frac{\theta\left(\eta\left(\tilde{e}_{t-1}/\bar{e}\right)^{\theta}-\gamma T\right)}{\gamma(1-\theta)}, \bar{e}\right] \tag{9.7e}$$

$$f_{t+1} = \left[\frac{(1-\alpha)\tilde{A}}{\tilde{w}_{t+1}}\right]^{\frac{1}{\alpha}}\frac{1}{\tilde{D}_{t+1}}. \tag{9.7f}$$

As in Chap. 7, the non-wage income from operating the family business raises fertility (9.7d). Traditional schooling is only different from the schooling in the modern sector if there are differences in initial conditions across the two sectors (9.7e). As in Chap. 8, the family farm provides income during retirement, a substitute for retirement saving when young (9.7c). Traditional households have lower wage income because of the constraints on their hours worked. Saving by traditional households will be relatively low because both their earnings and their saving rate are lower than households in the model sector. Thus, the traditional household will tend to have lower wage income and saving, and higher fertility, than households in the modern sector.

9.1.3 Labor Market Equilibrium and Locational Choice

We assume a national labor market so that the wages paid to a unit of human capital must be equalized across sectors, $\tilde{w}_t\tilde{D}_t = w_tD_t$. All productivity differences are "annual," based on differences in hours worked, and not due to differences in productivity per "hour." A national labor market requires that the total demand for labor from the traditional sector in each period t, $\frac{\tilde{N}_{t-1}}{\varphi_{t-1}\tilde{n}_{t-1}}f_t$, must exceed the supply of labor coming from younger members of traditional households in period t, $\tilde{N}_t\tilde{h}_t\tilde{z}\left(1-\eta\tilde{n}_{t+1}\right)+\tilde{N}_{t+1}\gamma\bar{h}\tilde{z}\left(T-\tilde{e}_t\right)$. Remember that each of the grandparents pass along the family technology to their children that remain in the traditional sector to share in their second period of life. So, if the population of older parents in the traditional sector is $\tilde{N}_{t-1}$, then the number of distinct farms is $\frac{\tilde{N}_{t-1}}{\varphi_{t-1}\tilde{n}_{t-1}}$, each of which demand f_t units of human capital from the younger generation of workers.

For some households to stay in the traditional sector and other households to leave the traditional sector, households must be indifferent about their locational choice. Equating the indirect utility functions for the generation-t traditional households under each option yields the following condition that determines the relative importance of income from the farm in equilibrium

$$\frac{O_{t+1} - w_{t+1}\tilde{D}_{t+1}f_{t+1}}{R_t w_t \tilde{D}_t \tilde{z}\tilde{h}_t \varphi_t \tilde{n}_t} = \Omega \equiv \left(\frac{1}{\tilde{z}}\right)^{\frac{1+\beta}{1+\beta+\psi}} - 1. \tag{9.8}$$

The condition indicates that the present value of *actual* relative income from the farm (the left-hand-side) must achieve a *required* value (the right-hand-side) in order to compensate traditional households for the hours-constraint on their work time when young. The lower is $\tilde{z}$, the greater must be the relative income from inheriting the farm. The degree to which residual income must compensate for the hours constraint is lessened by the exponent on the right-hand side, which is less than one. The exponent reflects the fact that, with endogenous fertility, being indifferent about working in the two sectors requires more than the equality of lifetime resources. Staying in the traditional sector to operate the farm also means having more children (higher wages when young raises the cost of children, while higher residual income when old does not). Given that parents like children ($\psi > 0$) the wealth compensation necessary for accepting less work hours when young is not as great when families are larger, as was first discussed in Chap. 7.

Condition (9.8) is important for all relationships in the model that involve the behavior and allocation of households across sectors. Condition (9.8) can be substituted back into (9.7a, 9.7b, 9.7c, 9.7d, 9.7e, 9.7f) to simplify the demand functions for traditional households.

The differences in saving behavior, fertility, and work effort across the households in the two sectors are all affected by Ω. Due to the presence of Ω, the saving rate out of wages is lower for traditional sector households, $(\beta - (1 + \psi)\Omega)/(1 + \beta + \psi) < \beta/(1 + \beta + \psi)$.

Comparing (9.7d) to (9.5d), shows fertility is higher in the traditional sector, even if schooling is the same across sectors. The ratio of traditional fertility to modern fertility is $1 + \Omega$.

Beyond the labor supply constraint that limits work for a traditional household relative to a modern household, the higher fertility of the traditional household will further reduce the adult equivalent units of labor supply (combining the labor of parents and children) relative to the modern household. This is because the increase in child labor does not fully offset the lost labor time associated with raising the child. The adult equivalent units of labor supply for a modern household can be shown to be $1 - n_{t+1}\left(\eta - \gamma\frac{\bar{h}}{h_t}(T - e_t)\right) = (1+\beta)/(1+\beta+\psi)$, while the supply from a traditional household is $(1+\beta-\psi\Omega)/(1+\beta+\psi)$, a lower value due to the presence of Ω.

Condition (9.8) can also be used to solve for the number of young households that must remain in the traditional sector. Using the first order conditions for labor demand in each sector, (9.8), and $\tilde{w}_t\tilde{D}_t = w_tD_t$, yields the following expression for the fraction of young households in period t that remain in the traditional sector,

$$\varphi_t = \frac{\left(\tilde{a}\frac{\tilde{D}_{t+1}}{D_{t+1}}\right)^{1/\alpha}(1+d)}{\Omega\tilde{n}_t(1-\alpha)Ak_t^{\alpha}\tilde{z}\tilde{h}_t}\left(\frac{1}{\tilde{D}_{t+1}}\right), \tag{9.9}$$

where $\tilde{a} \equiv \tilde{A}/A$ is the relative TFP in the traditional sector. Migration out of the traditional sector is greater the lower is the relative TFP value in the traditional sector and the higher is wage income. Migration is also affected by Ω because the higher is the required residual income ratio, the more siblings must exit the traditional sector. Technological progress, if balanced or favors the modern sector, also pulls labor out of the traditional sector.

Equation (9.9) determines the value of φ_t and the number of households in each sector. The sector-specific demographics are given by

$$\tilde{N}_t = \varphi_t\tilde{n}_t\tilde{N}_{t-1} \tag{9.10a}$$

$$N_t^* = n_tN_{t-1}^* + (1-\varphi_t)\tilde{n}_t\tilde{N}_{t-1}, \tag{9.10b}$$

where N_t^* denotes the number of young households in the modern sector during period t. We continue to use N_t to denote the total number of young households in the economy, i.e. $N_t = N_t^* + \tilde{N}_t$. Note that the total number of young households can also be written as $N_t = n_tN_{t-1}^* + \tilde{n}_t\tilde{N}_{t-1} = (n_t\pi_{t-1} + \tilde{n}_t(1-\pi_{t-1}))N_{t-1}$.

Using (9.10a) and (9.10b), the fraction of households living in the modern-sector can be defined recursively by the following difference equation

$$\pi_t = \frac{n_t\pi_{t-1} + (1-\varphi_t)\tilde{n}_t(1-\pi_{t-1})}{n_t\pi_{t-1} + \tilde{n}_t(1-\pi_{t-1})}. \tag{9.10c}$$

Note that π is no longer equivalent to the labor share in the modern sector, as it was in previous chapters. Here π is represents the fraction of young households that live in the modern sector. The labor share is now a more complicated concept and will be defined later. Note also that π will rise as φ_t falls, reflecting a smaller fraction of the children born in the traditional sector that remain there as adults. As indicated by (9.9), for φ_t to fall over time, requires growth in wages from physical and human capital accumulation and technological progress.

9.1.4 Capital Market Equilibrium

The capital supplied per unit of human capital employed in the modern sector is given by

$$k_{t+1} = \frac{s_t N_t + \tilde{s}_t \tilde{N}_t}{H_{t+1}}, \tag{9.11}$$

where

$$H_{t+1} = D_{t+1}\left\{ h_{t+1}\frac{1+\beta}{1+\beta+\psi}N_{t+1} + \tilde{h}_{t+1}\frac{1+\beta-\psi\Omega}{1+\beta+\psi}\tilde{N}_{t+1} - f_{t+1}\left(\frac{\tilde{N}_t}{\varphi_t \tilde{n}_t}\right)\right\}$$

is the total supply of human capital in the economy as a whole minus the human capital demanded to work in the traditional sector (remember that modern sector households can temporarily migrate to work in the traditional sector). Using the other equations of the model, (9.11) can be written out as the following transition equation for k,

$$k_{t+1} = \left[\overline{\beta}_t + \Omega\tilde{z}\left(\tilde{h}_t/h_t\right)(1-\pi_t)\right]\frac{(1-\alpha)Ak_t^{\alpha}}{1+g_t^{eff}} \tag{9.12}$$

where,

$$\overline{\beta}_t \equiv \frac{\beta\pi_t + (\beta-(1+\psi)\Omega)\tilde{z}\left(\tilde{h}_t/h_t\right)(1-\pi_t)}{1+\beta+\psi},$$

and

$$1+g_{t+1}^{eff} \equiv (n_{t+1}\pi_t + \tilde{n}_{t+1}(1-\pi_t))((1+\mathrm{d})h_{t+1}/h_t)$$
$$\left[\frac{1+\beta}{1+\beta+\psi}\pi_{t+1} + \frac{1+\beta-\psi\Omega}{1+\beta+\psi}\left(\tilde{z}\tilde{h}_{t+1}/h_{t+1}\right)(1-\pi_{t+1})\right].$$

The transition equation for physical capital intensity takes the same basic form as other we have encountered. Next period's physical capital intensity is determined by the average saving rate ($\overline{\beta}_t$) out of the current rental rate on human capital ($(1-\alpha)$ Ak_t^{α}) and by the growth rate of the effective labor supply (g_{t+1}^{eff}) that uses next period's capital stock.

The extension to a two-sector dual economy creates two new influences on physical capital intensity. First, the structural transformation releases labor from the traditional sector and increases the growth in the modern sector's working population. The size of the migration flow depends on how fast the traditional labor supply is growing relative to the traditional sector demand for labor. Faster growth in labor supply lowers capital intensity in the modern sector. This effect is captured by the expression in the denominator that describes the average work effort across sectors, $\frac{1+\beta}{1+\beta+\psi}\pi_{t+1} + \frac{1+\beta-\psi\Omega}{1+\beta+\psi}\left(\tilde{z}\tilde{h}_{t+1}/h_{t+1}\right)(1-\pi_{t+1})$. An increase in π increases the effective supply of labor because of an expansion in hours supplied and, possibly, an expansion in human capital per hour. For a given supply of labor, a decline in the traditional sector demand for labor also releases labor that must be absorbed in the modern sector. The demand for traditional labor is captured by the second term in the numerator (see *Problem* 2), $\Omega(1-\alpha)Ak_t^{\alpha}\tilde{z}\left(\tilde{h}_{t+1}/h_{t+1}\right)(1-\pi_t)$. As the traditional sector becomes smaller, this expression shrinks and the modern sector capital intensity falls.

The negative migration effect of the structural transformation on capital intensity is countered by a second effect, working through $\overline{\beta}_t$, that serves to increase capital intensity. Traditional sector households save at lower rates than modern sector households. As households migrate to the modern sector the economy's average saving rate increases and capital intensity rises. Overall, the structural transformation then has an ambiguous effect on physical capital intensity, just as we saw in Chap. 8.

9.2 Transitional Growth in the Long-Run

In this section we use the model to simulate growth paths. The simulations will reveal two characteristics that are not associated with standard one sector neoclassical models of physical capital accumulation. First, transitional growth is important for some time—even for centuries. Second the growth rates first rise and then stabilize for many decades before falling. As we shall discuss, these predictions of the model are consistent with the data. In contrast, the neoclassical model of physical capital accumulation predicts declining growth rates and a relatively quick convergence to the steady state. The predictions of the standard model have caused many to doubt the relevance of neoclassical growth, leading to the creation of *endogenous growth theory* that attempts to explain technological change or steady state growth. Our goal is to demonstrate that the transitional growth of an extended neoclassical theory can contribute significantly to our understanding of real world growth and development.

To begin, we further simplify the model in two ways. We assume that human capital is equal across the two sectors, i.e. the fundamental initial conditions for human capital are the same in both sectors. We also assume that exogenous technological change is balanced, so that the exogenous index of labor productivity grows at the same rate in each sector. We set $D_t = \tilde{D}_t$ because any level difference in the productivity differences can be captured by $\tilde{a} \equiv \tilde{A}/A$. These assumptions allow us to trace the dynamics of the economy using (9.5d), (9.5e), and the following three equations

$$\varphi_t = \frac{\tilde{a}^{1/\alpha}}{\Omega(1+\Omega)n_t(1-\alpha)Ak_t^\alpha \tilde{z}h_t}\left(\frac{1}{D_t}\right) \tag{9.13a}$$

$$\pi_t = \frac{\pi_{t-1} + (1-\varphi_t)(1+\Omega)(1-\pi_{t-1})}{\pi_{t-1} + (1+\Omega)(1-\pi_{t-1})} \tag{9.13b}$$

$$k_{t+1} = \frac{\beta(1-\alpha)Ak_t^\alpha[\pi_t + (1+\Omega)\tilde{z}(1-\pi_t)]}{n_{t+1}(\pi_t + (1+\Omega)(1-\pi_t))((1+\mathrm{d})h_{t+1}/h_t)[(1+\beta)\pi_{t+1} + (1+\beta-\psi\Omega)\tilde{z}(1-\pi_{t+1})]} \tag{9.13c}$$

Note that in (9.13c) that the right-hand side depends on k_{t+1} via π_{t+1}. This makes the computation a bit tricky because (9.13c) has to be solved implicitly for k_{t+1}.

In addition to (9.13a, 9.13b, and 9.13c), we will need to compute the average productivity per worker (*prod*) and the labor share in the traditional sector (*shr*)

$$prod_t = Ak_t^\alpha D_t h_t \frac{\frac{1+\beta}{1+\beta+\psi}\pi_t \quad + \quad \frac{1+\beta-\psi\Omega}{1+\beta+\psi}\tilde{z}(1-\pi_t)}{(1+n_t(T-e_t))\pi_t + \left(1+n_t(1+\Omega)\left(T-\tilde{e}_t\right)\right)(1-\pi_t)} \tag{9.13d}$$

$$shr_t = \frac{(1-\pi_{t-1})\Omega(1-\alpha)Ak_{t-1}^\alpha h_{t-1}\tilde{z}}{(1+d)n_t(\pi_{t-1} + (1+\Omega)(1-\pi_{t-1}))h_t k_t\left(\frac{1+\beta}{1+\beta+\psi}\pi_t + \frac{1+\beta-\psi\Omega}{1+\beta+\psi}\tilde{z}(1-\pi_t)\right)}. \tag{9.13e}$$

Equation (9.13d) gives worker productivity as measured by total output divided by the number of workers (young adults and their working-age children). Worker productivity grows as a result of the accumulation of physical and human capital as well as by an increase in the adult-equivalent labor supply per worker.

The contribution of the effective labor supply to worker productivity is captured by the last term which gives the adult equivalent labor supply divided by the number of workers. A decrease in the fraction of traditional sector households will cause worker productivity to rise because it increases work hours and lowers the number of child workers. Also a decline in fertility clearly raises labor supply per worker because it does not change the total labor supply coming from the family (reduced child labor is offset by higher adult labor one-for-one) but does reduce the number of family workers by lowering child labor.

Equation (9.13e) gives the share of labor in the traditional sector. It is the demand for human capital by the traditional sector divided by the total supply of human capital in the economy (see *Problem* 5).

9.2.1 Calibration and Numerical Analysis

We now analyze the model numerically. Many of the parameters are set using the calibration from the extended one-sector model in Chap. 5. The differences in calibration are based on the fact that it is no longer necessary to generate high fertility simply based on a low cost of children. We now have two household-types, one of which will have relatively high fertility because of the residual income from family production. As a consequence we set the parameter η, that determines the time cost of raising children, to a somewhat higher value than in Chap. 5 (0.18 rather than 0.165). This causes us to recalibrate some of the closely related parameters to maintain two other targets: a fertility rate of 1 and fulltime schooling in the steady state. The new parameters values are $\psi = 0.329$ and $\bar{e} = 0.10$.

Note also that there is a poverty trap under the calibration when $\gamma = 0.28$ (the value from Chap. 5). We assume that the economy overcomes this trap by invoking marginal increase in compulsory schooling or child labor legislation that lowers the work opportunity for children and γ. We first reduce γ just below 0.28 to initiate a relatively slow rise in schooling that we observe in countries that began modern growth around 1800. Later we consider more aggressive attempts to increase in schooling that are similar to what happen in the Asian Tiger Growth Miracle countries.[2]

As in Chap. 8, Table 8.2, we set $\tilde{z} = 2/3$ in order to match data on differences in annual hours worked across agriculture and manufacturing in the U.S. during the nineteenth century. This value implies that $\Omega = 0.394$, or that households in the traditional sector have about 40% more children than those in the modern sector. This number seems reasonable given that urban households had about 50% more children than rural households in the U.S. during the nineteenth century and that at least some of the fertility difference was likely due to small differences in schooling across rural and urban households (which we do not allow in the simulation model).

We set A and the initial value of D both equal to 1. The annualized growth rate of d is set to 1%, as in Chap. 5. We set the initial k to generate an initial annualized interest rate of 7%. Given the initial value of $k, \tilde{a}$ is set to target an initial value of φ of 0.90, using (9.13a), so that most traditional households remain in the traditional sector in the initial period. Finally, we choose initial values of $\pi = 1/3$ and $shr = 0.80$. Table 9.1 summarizes the parameter values of the calibration.

Before looking at the transition paths for the economy, first consider the consequences of the calibration for some of the individual components of (9.13c). The saving rate of traditional households is slightly negative, $\beta - (1 + \psi)\Omega = -0.0236$. So traditional households not only save less than modern households, but they are actually borrowers. An increase in π clearly raises the saving rate. However, recall that an increase in π also lowers physical capital intensity because it implies a reduction in the demand for labor in the traditional sector, which causes migration to the modern sector. The sign of the total effect that combines the saving effect with the demand for labor effect, is given by $1 - (1 + \Omega)\tilde{z} = 0.0704$. The saving effect slightly dominates the demand for labor effect, meaning the numerator of (9.13c) rises with π.

[2]In more realistic models of human capital accumulation the schooling take-off can be generated in ways other than through compulsory schooling or child labor laws. See Chaps. 4 and 5 for discussions of how schooling could be increased using subsidies, including tuition subsidies.

Table 9.1 Calibrated parameters

$\beta = 0.5000$
$\alpha = 0.3333$
$\gamma = 0.276$
$\eta = 0.1800$
$\theta = 0.4000$
$\psi = 0.3290$
$\tilde{z} = 0.6667$
$\Omega = 0.3940$
$\pi_0 = 0.3333$
$r_0^a = 0.070$
$A = 1$
$D_0 = 1, d^a = 0.01$

Now move to the denominator of (9.13c), where we can compute the expansion in the total labor supply associated with an increase in π. This effect is given by $\frac{1+\beta}{1+\beta+\psi}\left(1-\left(1-\frac{\psi\Omega}{1+\beta}\right)\tilde{z}\right) = 0.8201(0.3909) = 0.3206$. An increase in π increases the adult equivalent labor supply by almost one third of a full unit. Thus, the movement of households out of the traditional sector increases saving to fund investment but also creates a significant increase in the modern sector labor supply that must be supplied with physical capital.

Another effect on capital intensity comes through the effect of the structural transformation on the demographic transition. Migration of households out of the traditional sector will reduce the growth in the total population of young households and labor supply because households in the rural sector have 40% more children than households in the modern sector. The reduction in the population growth will make it easier for the economy to increase physical capital intensity. The question is whether the reduction in population growth will offset the release of labor, for a given population, when the traditional sector shrinks.

Table 9.2 gives the transitional growth over 10 periods (two centuries). Growth rates in output per worker rise over much of the first century, level off, and fall during the second century. There is no clear trend in growth rates over the two centuries. Interest rates also exhibit no clear downward trend until late in the second century. Conventional wisdom explains trendless growth rates and interest rates over long periods of time by an economy being close to its steady state. Indeed, the common observation of long-run trendless growth caused the profession to de-emphasize transitional growth in favor of developing *endogenous growth theory* that attempts to explain steady state growth caused by technological progress.

The problem with this conventional perspective is that in many respects real-world economies do not appear to be in a steady state. Despite trendless growth rates, the share of labor in manufacturing rises, fertility falls, and time investments in schooling increase—just as we see in Table 9.2. Robert Lucas (2002, p. 80), one of the founders of the endogenous growth approach, noted this inconsistency but decided to take an alternative approach in his analysis.

Table 9.2 Transitional growth and development

Period	0	1	2	3	4	5	6	7	8	9	10
$\%\Delta\ prod$ (annual)		1.06	1.65	1.92	1.99	1.95	1.86	1.79	1.74	1.68	1.60
shr	0.80	0.69	0.63	0.49	0.33	0.19	0.10	0.04	0.01	0.00	0.00
$r-\delta$ (annual)	0.070	0.078	0.081	0.082	0.081	0.078	0.073	0.067	0.061	0.055	0.050
n (per woman)	6.6	6.4	6.0	5.6	5.0	4.6	4.0	3.6	3.2	2.8	2.6
e	0.100	0.101	0.104	0.108	0.115	0.127	0.145	0.172	0.206	0.248	0.291

> One is left with two choices: first, we can identify increases in average schooling levels with net human capital investment. Since schooling levels are increasing in virtually all societies today, this is a possibility worth developing, but it cannot be pursued in a steady state framework. This is an important and neglected respect in which neither advanced nor most backward economies can be viewed as moving along balanced growth paths. Alternatively, we can think of a balanced path on which the time spent in schooling is constant but the quality of schooling is improving as a result of increases in general knowledge.

In this chapter we are making Lucas's first choice by demonstrating that trendless growth can be generated for long periods of time while an economy is in transition to a steady state.

The rising growth rates during the first century in Table 9.2 have little to do with capital accumulation. Interest rates are modestly increasing, indicating a drop in k, and schooling time rises only modestly. The rise in growth rates is primarily the result of the structural transformation and the associated rise in adult-equivalent labor supply, as workers work longer hours in the modern sector where they are unconstrained and where they have fewer children. At the end of the first century and beginning of the second century, physical and human capital accumulation takes over and keeps growth going, until diminishing returns begins to offset the rise in investment rates later in the century.

Rodrik (2013) sees similar growth patterns in the development of countries after WWII. He points out that the structural transformation can often generate fast growth during the early stages of development. This is particularly true when the structural transformation is from traditional agriculture to modern manufacturing, However, the growth associated with moving labor from low to high productivity sectors will a reach a limit. Additional growth must then result from the slow process of capital accumulation and technological progress across sectors.

Table 9.3 indicates that the most single important source of transitional growth after 200 years is an increase in the adult-equivalent labor supply. The increase in the adult-equivalent labor supply results from the migration of workers to the modern sector as well as the within-sector decline in fertility that occurs as schooling increases. More work in the modern sector and less child labor increase hours worked per worker. The endogenous transitional growth causes a 6.70-fold increase in worker productivity, slightly less than the growth due to exogenous technological change. After 200 years, technological change of 1% annually raises output per worker 7.32-fold. As we discuss later in this chapter, it is clear that we have formed a conservative estimate of the contribution from transitional growth. It is likely that the contribution of transitional growth significantly exceeds the technological progress that can be sustained in the long-run.

Table 9.3 Transitional growth

Growth due to	200 years of Growth
k	1.21
h	1.53
Labor supply	3.62
Total growth	6.70

Table 9.4 Rapid transitional growth

Period	0	1	2	3	4	5
%Δ *prod* (annual)		3.0	3.2	3.1	2.8	2.3
shr	0.80	0.70	0.56	0.37	0.21	0.09
$r - \delta$ (annual)	10.0	8.6	7.5	6.5	5.4	4.6
n (per woman)	6.6	5.2	4.0	3.0	2.6	2.2
e	0.10	0.13	0.18	0.25	0.33	0.41

Growth can be increased more dramatically when a less developed country takes a more aggressive approach to increasing schooling, as in the Asian Tiger countries. In Table 9.4, we consider a country that (i) has less capital intensity, an initial annual interest rate of 10% and (ii) pushes schooling more aggressively (reducing γ to 0.26). As a consequence, the growth rates are much higher. However, the pattern of growth is the same; rising growth rates that level off before declining.

9.3 Great Waves of Growth

Robert Gordon (1999) refers to U.S. growth as one "Big Wave." This is his way of describing the pattern where growth rates rise, level off for several decades or more, and then decline. Table 9.5 presents two measures of U.S. growth, GDP per capita and GDP per worker, for two centuries. Growth rates rose through much of the nineteenth century, leveled off for most of the twentieth century, and then declined in the last quarter of the century. Similar patterns can be observed in countries that began modern growth two centuries ago (see, for example, Farmer and Schelnast 2013, Figure 5.3). This growth rate pattern is similar to the growth pattern we simulated in Table 9.2.

One difference, between the simulation and U.S. historical growth, is the absence of a steady rise in the actual growth rate of worker productivity from the middle of the nineteenth century to the early decades of the twentieth century. As discussed in Chap. 8, hours worked per manufacturing worker fell during the nineteenth and early twentieth centuries, settling close to the now standard 40 h work-week before WWII. This pattern of hours worked is not captured in the model. The decline in hours worked over this period slowed the growth in output per worker in the data. Failing to capture the decline in hours-worked also means that the model overstates the contribution of increased labor supply to growth in worker productivity over the entire transition.

On the other hand, the model assumes no difference in worker productivity *per hour* worked across sectors. Accounting for some gap in productivity per hour worked, i.e. an hourly wage gap, would increase the contribution of the structural transformation to growth. This may not be a large omission for U.S. or European history, but could be in other settings where labor markets are less efficient. As documented by Gollin et al. (2014), after adjusting for differences in hours worked and education, the poorest developing countries today continue to exhibit large wage

Table 9.5 The Great wave: U.S. growth 1800–2000

Average annual growth rate in GDP per capita						
1800–1840	1840–1880	1880–1920	1920–1980	1980–1990	1990–2000	2000–2010
0.58	1.44	1.78	2.18	2.28	2.16	0.62
Average annual growth rate in GDP per worker						
1800–1840	1840–1880	1880–1920	1920–1980	1980–1990	1990–2000	2000–2010
1.07	1.58	1.47	1.92	1.58	1.62	

Notes: Average growth in GDP per capita is based on Farmer and Schelnast (2013. Table 5.1). Average growth in GDP per capita is from Chap. 2. Table 2.1

Table 9.6 Great wave: Asian tigers 1950–2010

Average annual growth rate in GDP per capita				
	1950–1965	1965–1980	1980–2000	2000–2010
Hong Kong	6.2	6.7	3.9	3.5
South Korea	3.4	7.5	7	3.8
Singapore	2.6	8.9	4.9	3.3
Taiwan	5.2	7.3	5.8	3.5

Notes: Average growth in GDP per capita is based on Farmer and Schelnast (2013. Table 5.1)

and productivity gaps. This is one of the reasons that McMillan and Rodrik (2011) find the structural transformation is so important in explaining the growth rate differences between Asia and Africa (see Chap. 7). In applying the model to the poorest currently developing countries, one would want to allow for a general wage gap that includes all sources of the difference in productivity across sector, similar to the approach we took in the first part of Chap. 7.

Table 9.6 reports the growth in GDP per capita in the Asian Tiger Growth Miracle countries. The more rapid and compressed growth we simulated in Table 9.4 is similar, although not nearly as rapid, as we see in the Asian Tigers. We discuss some of the reasons for the very rapid growth in these Growth Miracle countries in Sect. 9.4.

Overall, the general wave pattern is common to countries experiencing moderate growth for centuries and in countries experiencing more dramatic "Growth Miracles." Our model explains this growth pattern as follows. The diminishing returns associated with capital accumulation are initially more than overcome by the gains in productivity associated with the structural transformation and by rising investment rates. These two features of development cause growth rates to rise initially. However, there are ultimately growth slowdowns as investment rates level off and the structural transformation become complete.

The wave-like growth pattern offers a pessimistic forecast for growth in the twenty-first century. One can become even more pessimistic if there are reasons to believe that technological progress cannot continue indefinitely at the same rate we observed in the twentieth century. Jones (2002) extends the analysis of this chapter by relating technological progress to the growth in researchers (scientists and engineers engaged in research and development). In the twentieth century, the growth in researchers was based on population growth and on growth in research intensity (the fraction of the available work force devoted to research). Jones points out sustainable growth only comes from population growth (as with all investment rates, the fraction of the work force devoted to research is bounded). Assuming that population growth remains similar to that of the second half of the twentieth century, long-run growth is expected to be less than ½%.

The issue of twenty-first century growth was made popular by an article appearing in the *Economist* (January 12, 2013), entitled "Innovation Pessimism." The article suggests another reason to be pessimistic about growth. Academic research suggests that there may also be diminishing returns to research and development efforts (which Jones does not assume). For example, Vijg (2011) argues that the pace of technological progress will slow, and in fact has already begun to, particularly in the important areas of energy, transportation and medicine. This pessimism is contested by those who argue that the growth impact of innovations in computing, biotechnology, and personal communications has not yet been fully realized. Brynjolfsson and McAfee (2014) claim that we are just on the cusp of a second machine-age built around the computer. To maintain growth rates similar to the twentieth century, given the past importance of physical and human capital accumulation, it won't be enough to argue that technological progress will continue, it will have to accelerate. Given what we currently know, this seems unlikely.

9.4 South Korea: A Development Success Story

> The Korean experience suggests that mismanagement and not original position is more often the cause of *continuing* underdevelopment, combined on occasion with social and cultural attitudes—and not necessarily those associated with traditional values—held by authorities hostile toward the entrepreneurial and pragmatic spirit required for development. L.L. Wade and B.S. Kim, *Economic Development of South Korea* (p. vii)

At the end of the Korean conflict in 1953, the Korean peninsula was a war-ravaged, resource-scarce and over-populated backwater that was poorer than many countries in Africa or Latin America. The ceasefire provided a convenient natural experiment into the nature and causes of the wealth of nations. Korea was divided into a totalitarian, one party communist state in the north, the Democratic People's Republic of Korea (current population about 25 million). North Korea has languished in terms of economic attainment to the extent that it has required the mobilization of international food aid at times to avoid mass starvation of its citizens. In the south, the United Nations and United States helped establish and nourish the Republic of Korea, which has thrived. South Korea (current population 50 million) is the thirteenth most prosperous country in the world, with a per capita income higher than the European Union average. Today's Korea, centered on its vibrant capital Seoul, is a highly successful and innovative industrial country that organized a summer Olympics and is a member of the Organization of Economic Cooperation and Development (OEDC), the club of 34 advanced industrial economies based in Paris, France.

At independence, the overwhelming majority of Korea's resources and industry were located in the north. South Korea had been devastated by the war, and its refugee-swelled population had to depend on basic agriculture and US food aid for survival. Yet, while its initial industrial and resource bases were thin, South Korea was able during the following 60 years to register an impressive economic record, including economic growth that averages over 7% per annum in real terms. South Koreans invested heavily in human and physical capital and also adopted an economic model that combines an outward-oriented development strategy with state direction of the financial system and the real economy. Hard work, innovation, trade openness and the promotion of exports through trade-friendly policies, such as realistic exchange rates and reduction of tariffs and non-tariff trade barriers, have made South Korea a global economic power house.

9.4.1 The Onset of Growth

Several preconditions necessary for the subsequent growth take off were established during the 1950s. First, a land reform was enacted in 1949, distributing to peasants lands that became available following the end of Japanese rule in 1945. This was instrumental in establishing property rights and in pacifying rural areas. It also weakened the political power of large landlords that tends to inhibit pro-growth policy. Second, grass root demand for education exploded, reflecting Korean family and societal values and repression of the Korean culture during the long foreign dominations. The education drive laid the foundation for later achievements as it provided younger generations with the skills they would later need in the subsequent industrialization drive. Despite these early achievements, however, population growth in the 1950s was such that per capita incomes did not rise, a coherent long-term economic development strategy was lacking, and growth in part reflected temporary US relief and reconstruction assistance.

9.4.2 Growth Acceleration

South Korea's growth surge began in earnest in 1963. Elections that year followed political unrest and a military intervention in 1961. In the years that followed, the political system was stable and the government was able to formulate effective investment plans without political interference, which succeeded in fostering economic development and helped it gain greater political legitimacy. South Korea's rate of economic growth between 1965 and 1972 was truly impressive, over 8% per annum, second only to Singapore's and Japan's. During this period an impressive shift occurred in the composition of GDP away from agriculture toward industry.

The share of agriculture declined from about 40% to about 25%, while that of manufacturing rose from the high teens to about 30%. Also rising were the shares in GDP accounted for by construction, industry-related services, finance (banking and insurance) and education. Regarding the size of government employment, whose swelling in developing countries is often cited as a factor inhibiting development, Korea offers an interesting lesson: during this period, there was a decline in the size of Korea's government (including public administration and defense).

The expansion of Korea's manufacturing sector was broad-based, including the production of transportation equipment (cars, ship building), electrical machinery and chemicals and petroleum products. This broad-based industrial development effort was made possible by a high saving rate, which fostered physical capital accumulation, and also by Korean families' legendary drive for educational attainment, as human capital investments led to the improved workforce skills. Abundance of physical and human capital went hand in hand with investments in research and development (R&D), which facilitated imitation and innovation—the adoption of existing technologies and the creation of new ones. Korea also imported large amounts of capital from abroad, which embodied newer technologies and enabled South Korean conglomerates to export their manufactures abroad.

Successful Korean firms (Samsung, Hyundai) have become familiar global brands that earn and invest in R&D billions of dollars every year. They are the most visible of a number of diversified conglomerates called *chaebol*, which facilitated Korea's industrial prowess. Controlled by powerful families, these conglomerates' success lies in a combination of visionary management, special advantages enjoyed in the domestic market place, and also adherence to market discipline. Among the advantages, preferential access of the chaebol to affordable finance from Korea's banking system was crucial for their expansion and competitive advantage internationally. This is a form of "industrial policy," where government policy favors certain industries. However, the market discipline for the chaebol was provided by their export orientation: firms able to survive in the international market place were promoted, which helped steer the economy to internationally profitable sectors.

9.4.3 Lessons

What are some of the lessons behind South Korea's remarkable transformation from a poor, resource-scarce, over-populated backwater to an advanced country in the span of 60 years (two generations)? One takeaway concerns the role of national will and determination in overcoming obstacles to development—recall the discussion of the role of ownership in Chap. 3. In 1950, Korea was a poor but classless society. One of the early lessons all Koreans internalized was that foreign aid was transitory and that their future depended on their actions. Government and private sector joined forces on a path to capital accumulation (broadly construed to include human and physical capital).

There were important sources of growth not captured by the model from Sect. 7.2. These omissions explain why our model could not generate the very high growth associated with the Asian economic miracles. First, the South Korean government invested heavily in public infrastructure, which was modeled in Chaps. 3 and 5, but omitted for simplicity in Chap. 9. Second, the government used an industrial policy that subsidized private capital formation. Industrial policy in promoting private capital formation is a common but controversial feature of the Asian Tigers approach to growth. Finally, the Korean economy was opened to international trade and capital flows. We saw in Chap. 5 that opening developing economies to international capital flows have the potential to accelerate growth. International trade can also increase growth by increasing the pace of the structural transformation for countries with a comparative advantage in manufacturing, as we have discussed in Chaps. 6, 7, and 8. Furthermore, the international competition associated with an open economy helped discipline Korean industries supported by the government. The government subsidized those industries that performed best in the open economy. For calibration studies of South Korea's Growth Miracle, including open economy models, see Connolly and Yi (2015) and the references they cite.

Korea's exciting economic story has been accompanied by an equally impressive political evolution. Since the 1980s, it has evolved into an open, free, and thriving democracy. While growth can be initiated by pro-growth dictatorships, the rise of democratic institutions helps limit the rise of government corruption and mismanagement that could derail growth.

9.4.4 Challenges

While Korea's broad-based success and remarkable decline in poverty have helped maintain social cohesion, challenges remain. One important challenge is to address widening income inequality. Like other countries, there are two labor market tracks in Korea: many workers have regular, stable jobs that pay decent wages and benefits, while others (about a third of the total) are only able to get temporary or part time employment. To address such *labor market duality* and make labor markets more inclusive, it is important to expand training and facilitate access to job opportunities of certain groups.

The participation of young Koreans in the labor force is lower than the OECD average, reflecting in part high enrollments in tertiary education, but also lack of skills demanded by firms, which could be addressed by expanding training. The participation of women in the labor force is much less than that of men, which offers a big opportunity going forward: as Korea ages—it will be one of the oldest societies in the world by the middle of the century—more women holding paid jobs could help raise Korea's growth.

There is also room to improve the productivity of firms in non-traded goods sectors, including many services, where firms are small and wages are lower than in the export-oriented sectors. Another challenge comes from the power of Korea's chaebol and other corporate groups and the challenge they pose for regulators.

Longer term challenges, as already alluded to, also arise from Korea's rapidly aging population. Finally, Korea faces the prospect of reunification and absorbing into its thriving economy the 25 million Koreans living in impoverished North Korea.

9.5 Human Capital Extensions

There are various ways that one can extend the human capital production function in order to explain more of the growth in worker productivity. In our simple model, we include only student time spent in school. Other inputs are also important, including goods and service inputs (e.g. tuition costs that include teachers, books, and other classroom materials), the human capital of people in the community (including parents, relatives, and teachers), and health inputs that improve student learning and worker productivity directly.

9.5.1 Goods Inputs

We mentioned in earlier chapters that purchased goods and services associated with schooling are needed to develop human capital. Some argue that variations in goods inputs have little effect on human capital and therefore can be reasonably omitted. However, this is far from clear as the evidence on the effect of "school quality" on human capital formation is mixed.

For example, Krueger (2002) estimates that an increase in school quality measures (such as class size) not only increases human capital, but they do so at a reasonably high 6% rate of return. Even if *marginal* changes in school quality have small effects, at *current levels of school quality in developed countries*, this does not imply that improvements in school quality are unimportant when the level of overall quality is very low, as in developing countries. In fact, there are several reasons to expect that marginal changes in quality could have high payoffs at low levels while having low payoffs at high levels.

First, there are surely diminishing returns to reducing class size and increasing the availability of textbooks and other school inputs. Second, there are likely to be compositional differences in school expenditures in poor and rich countries. In rich countries increases in school expenditures could reflect a variety of factors, some of which are likely not closely related to learning. In poor countries, increased expenditures are more likely to reflect increases in fundamental inputs such as teachers and textbooks. Third, school expenditures in poor countries are more concentrated on young children, since few older children are in school. Evidence suggests that the returns to education investments of all types are highest for young children (Heckman 1999, Carneiro and Heckman 2003).

9.5.2 Community Externalities

There is little evidence of significant human capital externalities at the state, national, or international levels (Aghion and Howitt 1998, Chap. 11; Heckman and Klenow 1998; Acemoglu and Angrist 1999; Krueger and Lindahl 2001; Ciccone and Peri 2006). However, Borjas (1992, 1995) finds important intergenerational human capital spillovers at the community level in United States data. This is impressive since, as with school quality, one would expect there to be diminishing returns to human capital externalities (Heckman 1999). When education levels are generally low, there might be a much greater impact on children's learning from increasing the average education level of teachers, parents, and role models in the community than when the education levels are high. This makes it harder to identify external effects from data collected in countries where the population is relatively well-educated. To this point, there is consistent evidence of community-level externalities in developing countries (Shavit and Pierce 1991; La Ferrara 2003; Cox and Fafchamps 2007; Angelucci et al. 2010; Wantchekon et al. 2013).

Extending the human capital production function to include goods inputs and community externalities could take the form, $h_{t+1} = x_t^{\theta_1} e_t^{\theta_2} h_t^{\theta_3}$, where x is the goods input and where the community externality is captured by the human capital of the previous generation. As long as the three exponents still sum to be less than one, we have diminishing returns and can explain growth rate slowdowns. Adding goods inputs and community externalities will increase the contribution of human capital to explaining transitional growth. In addition, transitional growth will extend, at higher rates, over longer period before the slowdown begins. Rangazas (2002, 2005) does a quantitative analysis of growth with similar human capital production functions and shows significant increases in the duration of transitional growth.

9.5.3 Health Investments

We should at least mention the growing literature on how health interacts with learning and worker productivity. First, there is evidence that nutrition directly affects worker productivity in currently developing countries (Strauss and Thomas 1998) and in historical settings (Fogel 1997). Second, early (Bharadwaj et al. 2013) and contemporaneous (Miguel and Kremer 2004; Bleakley 2007) health interventions in children improve educational outcomes. Finally, there is at least some evidence that increases in longevity encourages individuals to invest in more schooling because the expected payoff period increases (Oster et al. 2013).

9.6 Convergence Revisited

The neoclassical growth models we have used predict that if the fundamentals of countries are similar then, regardless of when modern growth begins, all countries should eventually converge to similar long-run per capita income levels and living standards. Fundamentals refer to the preferences, technology, and policies of the country. The convergence prediction requires that countries that begin modern growth later must grow *faster* than countries that began modern growth earlier, in order to "catch-up."

In the standard neoclassical growth model of Chap. 2, economic growth rates uniformly fall with the level of income as countries develop, making the prediction of convergence obvious and potentially rapid. However, the growth model of this chapter suggests that convergence may be a very slow process. From Table 9.2, we see that growth rates first rise and then exhibit little trend for many periods. In this model, a country that begins modern growth later will typically first grow *slower than*, and then grow at a rate *similar to*, the leaders in per capita income. There could easily be no sign of convergence for two centuries after growth first begins in the lagging country. This may be true even if the fundamentals of the leading and lagging economies are similar.

We noted in Table 9.5 that the pattern of rising and then moderate and trendless growth rates over long periods of time was evident in U.S. history. The same was true for the history of all currently high income countries, with the exception of a relatively small number of Growth Miracle countries. Now consider the currently developing countries, whose growth take-off lagged that of the current leaders. Since World War II, most countries of the world have begun sustained economic growth. The standard neoclassical growth model from Chap. 2 would predict that many of these countries, those that have established decent fundamentals, would grow faster than the leaders. However, the average growth rates of countries at every stage of development has been similar to that of the income leaders (see, for example, Jones and Vollrath 2013, Figure 3.6 and Kraay and McKenzie 2014, Table 9.1). This means that, on average, we have not seen countries converge. The lack of convergence for many decades, while not consistent with the standard neoclassical model, is quite consistent with extended neoclassical model from this chapter.

A recent study by Im and Rosenblatt (2013) looked at the lack of convergence in detail. The main question was whether developing countries have grown faster (in per capita terms) than the rich countries, the requirement for convergence. The study used the average Post WWII growth rates of high-income countries as the benchmark. Over the 50-year period 1961–2011, high income countries grew about 2% per year. If one looks at the average per capita growth rates for 141 developing countries over this period, the average growth rate was also about 2%. Thus, on average, developing countries are not converging. More precisely, 80 of the developing countries grew by at least 1.5%—of which, 64 grew by at least 2%. While these 64 countries are converging, for most the convergence is painfully slow. Thirty-one countries grew by 3% or more, while only 9 reached 5% growth. So, noticeable convergence is only taking place in a minority of developing countries. It is even fewer developing countries that have experienced Growth Miracles and rapid convergence.

There has been some concern that the Growth Miracles countries have recently experienced a decrease in growth rates, slowing their convergence. Evidence of this can be seen in Table 9.6. As we have discussed, our model predicts that the high growth rates of countries that generate unusually high levels of investment in physical and human capital, coupled with unusually rapid structural transformations, will naturally experience growth slowdowns, as exhibited in Table 9.4. It is not clear that special theories of "middle-income traps" are needed to explain these slowdowns (see, for example, Agenor and Canuto 2012, and the Aiyar et al. 2013). The middle-income trap theories focus on the possibility that TFP growth may slow-down in countries that are not able to compete in the production of human-capital-intensive goods associated with a fast pace of innovation. However, TFP growth will slow for more basic reasons such as slowdowns in the efficiency gains associated with the structural transformation or in the quality dimensions of schooling that are not typically measured (e.g. length of the school year, teacher inputs, the average ability of students attending school). Nor should there be major concerns about most Growth Miracle countries since, even at their slower growth rates, they are converging quite rapidly to the income leaders of the world.

Im and Rosenblatt (2013) looked for evidence of "middle-income traps" in the form of a general tendency of middle-income countries to experience abrupt slowdowns or other unusual growth patterns. They found that the pattern of growth was no different for middle-income countries than for countries at other income levels. Middle-income countries that have experienced Growth Miracles, or even above average growth, are naturally going to experience more dramatic growth slowdowns than those in moderately growing economies. However, this does not seem to be the result of a special trap that afflicts all middle-income countries. The average middle-income country is experiencing similar growth rates, and the associated very slow convergence, that the average country at *all* levels of income is experiencing in the Post WWII period.

We discussed in Chap. 7 that the growth experience of African countries has been particularly disappointing. In recent years there has been some renewed hope that Africa is finally beginning to convergence. From 2000 to 2011, per capita income in Africa grew 4.7% annually (Devarajan and Fengler 2012). The increase in growth rates is attributed to several factors including improved macroeconomic policies and the early stages of a demographic transition. Poor macroeconomic policies have been shown to create economic disasters that have long-term negative consequences for growth (such as hyper-inflations). The elimination of bad macroeconomic policies, give growth a chance (Easterly 2005). Africa is also benefiting from a demographic dividend associated with falling fertility. The most rapidly growing demographic group in Africa is now working-age adults (Devarajan and Fengler 2012, pp. 13–14). The number of adult workers is increasing rapidly relative to the number of children, causing worker productivity to rise for many of the reasons we have stressed in the book.

Despite the recent success, there are doubts about whether Africa's growth can be sustained (Rodrik 2011; Devarajan and Fengler 2012). The concerns center around the weak structural transformation in Africa. Most workers are still employed in traditional agriculture and small informal family businesses. For example, 70% of the work force in Uganda is employed in agriculture and informal family businesses.

As mentioned in Chap. 7, much of Africa's recent success has been caused by a boom in international commodity prices that while benefiting some groups in the economy also creates a de-industrialization that slows the structural transformation and sends labor into less productive occupations. This has caused some economists to support industrial policies, similar to those that sped the structural transformation in Asia, to raise employment in the formal manufacturing sector where labor productivity is higher (see Sect. 9.4, for example).

There are domestic sources of the weak structural transformation in Africa as well. The continent is lacking in public infrastructure. Many of the services associated with the public infrastructure that does exist are poorly provided. Ports are inefficiently run, creating long delays in the shipping and distribution of products. Utilities are unevenly distributed according to political patronage. The trucking industry in many African countries is protected and operates with very high profit margins, driving up transportation costs. Regulatory costs associated with starting and running businesses are the highest in the world. The quality of schools is very poor due to poorly educated and frequently absent teachers. All of these features drive business costs up and labor productivity down in the formal manufacturing sector.

9.7 Politics and Growth

We have seen that the Great Divergence is a byproduct of the timing and nature of modern growth. Once growth begins, there is no strong force that slows growth down, making it difficult for lagging countries to catch up to leading countries. The slow pace of the structural transformation and the rising investment rates in physical and human capital per worker keep growth rates from falling over a lengthy transition marked by relatively trendless growth rates. For a lagging country to converge, requires its government to take unusual actions that accelerate the structural transformation and the rise in investment rates beyond the historical norm of the leaders. Only a few countries have been able to do this to the extent required for a rapid convergence to the leaders' standard of living.

All of this means that the timing of the initial take-off of sustained modern growth is generally the key to explaining the differences in the standard of living across countries today. The take-off of the modern sector means conditions favor production that is intensive in physical and general human capital over traditional methods of production that rely heavily on land, raw labor, and the specific human capital that is passed down over many generations of family-based production. This raises the question of what prevents the conditions needed for the modern sector to begin to dominate over traditional production (see Chap. 6 for a precise statement of the conditions).

Acemoglu and Robinson (2012) offer an explanation of why the take-off of modern growth is delayed in many countries. They argue for a type of poverty trap that becomes engrained in the political institutions of a country. When a country is undeveloped it is nevertheless possible for a small fraction of the population to become rich by exploiting the available resources and labor. The rich elite are naturally also those that have seized political power. The political power allows the elite to create policies and laws that maintain the status quo and prevent economic competition from anyone outside the small fraction of the population that rules the country. Monarchies, dictators, one-party rulers, and overly influential landowners and traditional craftsmen maintain control by creating monopolies, blocking innovation, and arbitrarily seizing land and taxing income. For our purposes, one can think of this situation as one where the traditional sector of the economy is protected from competition because most of the income generated there is flowing to a small group in power.

To break this type of poverty trap requires a major change in political institutions. Changing political institutions under an autocratic regime requires conflict. For example, sustained modern growth is generally believed to have started with the Industrial Revolution in England. Acemoglu and Robinson argue that the Industrial Revolution never would have occurred without a long political struggle, involving much unrest and even civil war, culminating in the Glorious Revolution. The political struggle shifted power from the autocratic monarchy to the more democratic Parliament. The Parliament, representing a broad coalition of interest throughout the country, ended the creation of monopolies, the selective granting of patents, and the arbitrary taxation and seizures of property that were the hallmarks of the monarchy's reign. In addition, the government became more focused on the general infrastructure of the economy; roads, banks, and education for the masses. The country became ruled by more even-handed laws and policies and not by a ruling elite that intended to maintain its power and wealth. The dramatic change from an exclusive and extractive political rule to a inclusive one, leveled the playing field and set the stage for the Industrial Revolution.

9.7.1 Natural Resources and the Politics Trap

In Chaps. 6 and 8 we discussed how natural resources can potentially slow growth by creating comparative advantages in the traditional agriculture and commodity sector. Especially in open economies, a comparative advantage in the traditional sector slows the structural transformation and adds volatility to the economy that slows growth. The natural resources of a country also play a role in the political poverty trap by weakening democracies and inducing the creation of autocratic regimes that are looking to get rich quick by gaining control of the income flows from the traditional sector (Collier 2007, Chaps. 2 and 3).

Ironically, autocratic regimes can also create Growth Miracles (Glaeser et al. 2004). After WWII, the Asian Tigers began as dictatorships. However, they were pro-growth dictatorships, imposing policies that forced and encouraged public and private investments of various types on their poor populations. In democratic regimes, with limited central government power, high levels of saving and investment are not going to be the preferred choice of an impoverished electorate. So, when do autocratic governments work for growth and when don't they?

Autocracies are clearly risky for the growth; they can lead to Growth Miracles or Growth Disasters. It appears that autocracies are particularly detrimental for growth when the country's population is ethnically diverse (Collier 2007, Chap. 3). In this case, the tendency of those in power is to control resources for its own ethnic group at the expense of the country as a whole. A particularly bad mix for a developing country is to have an autocratic regime, with an ethnically diverse population, and rich natural resources. Unfortunately, this toxic mix is quite common in developing countries, particularly in Africa and the Middle East. The relatively low current standard of living in these underdeveloped regions is largely the result of a legacy of extractive regimes that have been generated by the combination of autocratic government, ethnic diversity, and natural resources.

9.8 The Structural Transformation in Later Stages of Development

Throughout the book our focus has been on the early structural transformation, characterized by a reallocation of labor from traditional agricultural and informal shops to manufacturing in modern firms. In the later stages of development, the structural transformation continues but in a quite different fashion. Manufacturing begins to shrink and the formal service industry grows into the dominant sector for employment. This section offers a taste of the research and issues related to the late structural transformation from a manufacturing economy to a service economy.

9.8.1 Convergence Once Again

Duarte and Restuccia (2010) study the relative convergence of middle and high income countries to the United States during the Post WWII period. They seek to explain why many countries converged in the early part of the Post War period, only to see their convergence stall out or even regress over more recent decades.

Their explanation is based on the common pattern of labor allocation that occurs over the long course of development: employment shifts from agriculture to manufacturing and then to services. They combine this labor allocation pattern with an observation about the relative productivity of workers across countries. Workers from middle and high income countries have similar labor productivity to the United States in manufacturing but have relatively low productivity in both agriculture and services.

The implication of the common pattern of labor allocation over the course of development, along with the observation about the relative cross-country productivity of workers from different sectors, helps explain why convergence occurs and then stalls. In early development, the average worker productivity in lagging countries converges to the United States as workers move out of relatively low productivity agriculture to manufacturing where worker productivity is on par with the United States. As development continues, workers begin leaving manufacturing for the service sector where labor has low productivity relative to the United States. The late structural transformation from manufacturing to services causes the relative convergence in average labor productivity to stall.

9.8.2 Why Does the United States Work More?

Rogerson (2008) studies hours worked across countries rather than labor productivity. He attempts to explain the decline in the total hours worked of the adult aged population (ages 15 to 64) in Europe relative to the United States. Similar to Duarte and Restuccia (2010), the explanation is closely tied to differences in the service sector across countries.

Rogerson documents that tax rates have increased in Europe relative to the United States in the Post WWII period. High tax rates encourage a substitution of taxed activity for untaxed activity that generates similar utility. Rather than working and purchasing in the market, one can avoid taxes by producing similar goods at home. The area where home production is the closest substitute for market production is service provision: watching television instead of going to the movie theater, making meals at home instead of eating in restaurants, using your own car instead of cabs, trains, and planes.

Rogerson's explanation for the greater market work effort in the United States uses these components to solve the cross-country work puzzle. High taxes in Europe causes households to substitute away from working and producing in the market service sector—a major employer of labor in advanced economies in the late stages of development. As a consequence, recorded hours worked in Europe lags market work in the relatively low-tax United States where more services are demanded from the market.

9.8.3 Premature Deindustrialization

Rodrik (2015) is concerned about sustaining growth in today's developing countries. He sees a tendency for growth in poor and middle income countries to slow, preventing the much-needed rise in living standards. As with the previous authors of this section, his analysis begins with the historical patterns of the structural transformation. For currently rich countries, we have seen the relative decline in the manufacturing sector after WWII. Right after the War, the United States and Great Britain, for example, had between one quarter and one third of their labor allocated to manufacturing. Currently their labor shares in manufacturing are about 10%. Rodrik attributes the late development pattern in labor allocation to labor-saving technological progress (automation). The workers replaced by machines in manufacturing are needed to operate and maintain the equipment. Many of these new occupations are classified as being in the service sector.

Rodrik further notes that the deindustrialization of the work force has begun earlier in development, i.e. at lower income levels, in currently developing countries.

He terms this phenomenon, "premature deindustrialization." The reasons for and consequences of *premature* deindustrialization are much different than those resulting from labor-saving automation in manufacturing, as we have already discussed in Chaps. 7 and 8. Recall that a major reason for deindustrialization in developing countries of the late nineteenth and late twentieth centuries was the globalization of trade in those periods. Countries lacking a comparative advantage in manufacturing will see that sector contract when their economies are opened to trade. The trade-induced reversal in the normal pattern of the structural transformation is potentially harmful to aggregate growth.

Growth is lowered by a stunted structural transformation if there are large productivity gaps across traditional and modern modes of production. These gaps are more likely to exist the stronger are the "ties that bind" workers to the traditional sector, as discussed in Chap. 7. If wages are prevented from rising in the modern manufacturing sector, the ties to the traditional sector will not be broken enough to generate a productivity enhancing reallocation of labor. Workers will forgo more productive work opportunities to remain in agriculture and the informal service sector because of the chance of inheriting the family farm or business and because of the informal insurance provided in small communities.

9.9 Conclusion

In this chapter we use a complete dual economy model to conduct a quantitative study of transitional growth over long periods of time. Unlike the standard neoclassical growth model, that predicts declining growth rates and rapid convergence to a steady state, the dual economy generates rising growth rates that eventually level off for several decades before finally begin to decline after a long transition.

The analysis shows that the early stages of growth are dominated by the structural transformation. The migration of labor from the traditional sector to the modern sector expands work hours and reduces the number of child workers, resulting in a rise in productivity per worker. Later in the growth process, we see rising investment rates in physical and human capital that sustain high growth rates for many decades. Finally, as the rise in investment rates stabilize, diminishing returns dominates, and growth rates begin to fall.

We left out several factors that would increase the ability to explain growth and extend the duration of the transition with growth rates. These factors include broader investments in human capital, human capital externalities, and public investments in infrastructure and R&D. Eventually all of these sources of growth will weaken and growth rates will decline. This suggests that growth in the twenty-first century will likely be weaker than in the twentieth century.

9.10 Exercises

Questions

1. Explain why young households may choose to remain in the traditional sector even if they do not inherit land from their parents or village.
2. In the model used in this chapter, how do traditional and modern households differ in their behavior and why?
3. What does it mean for there to be a "national" labor market? Does it imply that annual wages are the same for modern and traditional workers? What condition must be satisfied for the labor market to be national?
4. What is the "required" residual income from operating a family business? What would make the "required" income increase? How does an increase in required income affect
   (a) the saving rate of traditional households
   (b) fertility of a traditional household
   (c) family labor supply of a traditional household in adult-equivalent units
   (d) fraction of households that live in the traditional sector
5. How does the structural transformation affect the capital-labor ratio in the modern sector? The rental rate on human capital throughout the economy?
6. List all the ways that the structural transformation may affect economic growth.
7. What features of economic growth do the simulations reported in Tables 9.2, 9.3, and 9.4 capture well and what features do they miss?
8. Based on Table 9.3, what is the most important quantitative connection between the structural transformation and growth in worker productivity? Explain the connection in detail.
9. Why can the model of this chapter be viewed as generating a conservative estimate of the role of endogenous transitional growth in explaining real world growth?
10. What are the implications for twenty-first century growth that are suggested by the research discussed in this chapter?
11. Discuss the events that initiated the South Korean Growth Miracle.
12. Why does the standard growth model from Chap. 2 have trouble explaining the lack of convergence across economies of the world? How does the extended model of this chapter help explain the lack of convergence?
13. What has slowed the structural transformation in Africa?
14. Explain the political poverty trap.

Problems

1. Use the first order conditions for labor demand in each sector, (9.8), and $\tilde{w}_t\tilde{D}_t = w_t D_t$, to derive (9.9), the fraction of young households in period t that remain in the traditional sector.
2. Show that the total demand for labor from the traditional sector can be written as $\tilde{N}_t \dfrac{\Omega(1-\alpha)Ak_t^{\alpha}\tilde{z}\tilde{h}_t}{(1+d)k_{t+1}}$.

3. Derive (9.12) and show that it simplifies to become (9.13c). Explain how the structural transformation affects physical capital intensity.
4. Carefully explain why one cannot explicitly solve (9.13c) for k_{t+1}.
5. Derive (9.13e). Explain the different ways that the structural transformation affects the traditional labor share.

References

Acemoglu D, Angrist J (1999) How large are the social returns to education? Evidence from compulsory schooling laws. NBER working paper 7444. National Bureau of Economic Research, Cambridge, Massachusetts.

Acemoglu D, Robinson J (2012) Why nations fail. The origins of power, prosperity, and poverty. Crown Publishers, New York.

Agenor P, Canuto O (2012) Middle-income growth traps. World Bank policy research working paper 6210. World Bank, Washington, DC.

Aghion P, Howitt P (1998) Endogenous growth theory. MIT Press, Cambridge, Massachusetts.

Aiyar S, Duval R, Puy D, Wu Y, Zhang L (2013) Growth slowdowns and the middle-income trap. IMF working paper 13/71. International Monetary Fund, Washington, DC.

Angelucci M, De Giorgi G, Rangel M, Rasul I (2010) Family networks and school enrollment: Evidence from a randomized social experiment. Journal of public economics 94(3–4):197–221.

Banerjee A, Duflo E (2011) Poor economics. Public Affairs, New York.

Bharadwaj P, Loken K, Neilson C (2013) Early life interventions and academic achievement. American economic review 103(5):1862–1891.

Bleakley H (2007) Disease and development: Evidence from hookworm eradication in the American south. Quarterly journal of economics 122(1):73–117.

Borjas GJ (1992) Ethnic capital and intergenerational mobility. Quarterly journal of economics 107 (1):123–150.

Borjas GJ (1995) Ethnicity, neighborhoods, and human-capital externalities. American economic review 85(3):365–390.

Brynjolfsson E, McAfee A (2014) The second-machine age. W.W. Norton, New York.

Carneiro P, Heckman J (2003) Human capital policy. In: Heckman J, Krueger A, Friedman B (eds) Inequality in America: What role for human capital policies? MIT Press, Cambridge, Massachusetts.

Carter SB, Ransom RL, Sutch R (2003) Family matters: The life-cycle transition and the unparalleled fertility decline in antebellum America. In: Guinnane TW, Sundstrom WA, Whately W (eds) History matters: Essays on economic growth, technology, and demographic change. Stanford University Press, Stanford, California.

Ciccone A, Peri G (2006) Identifying human capital externalities: Theory with applications. Review of economic studies 73(2):381–412.

Collier P 2007 The bottom billion. Oxford University Press, Oxford.

Connolly M. and Yi, K (2015) How Much of South Korea's Growth Miracle can be Explained by Trade Policy? American Economic Journal: Macroeconomics 7(4) 188–221.

Cox D, Fafchamps M (2007) Extended family and kinship networks: Economic insights and evolutionary directions. In: Schultz P, Strauss JA (eds) Handbook of development economics, vol. 4. Elsevier North-Holland, Amsterdam, pp 3711–3784.

Devarajan S, Fengler W (2012) Is Africa's recent economic growth sustainable? Institut Francais des Relations Internationals, Paris and Brussels.

Duarte, M. and Restuccia, D., 2010, The Role of the Structural Transformation in Aggregate Productivity, *Quarterly Journal of Econoimics*, pp.129–173.

Easterly W (2005) National policies and economic growth: A reappraisal. In: Agion P, Durlauf S (eds) Handbook of economic growth, vol. 1A. Elsevier North-Holland, Amsterdam.

Farmer K, Schelnast M (2013) Growth and international trade: An introduction to the overlapping generations approach. Spring-Verlag, Berlin.

Fogel R (1997) New finding on secular trends in nutrition and mortality: Some implications for population theory. In: Rosenzweig M. Stark O (eds) Handbook of population and family economics. North Holland, Amsterdam, pp 433–481.

Glaeser E, La Porta R, Lopez-de-Salines F, Shleifer A (2004) Do institutions cause growth? Journal of economic growth 9(3):271–303.

Gollin D, Lagakos D, Waugh M (2014) The agricultural productivity gap. Quarterly journal of economics 129(2):939–993.

Gordon R (1999) U.S. economic growth since 1870: One big wave? American economic review 89 (2):123–128.

Heckman J (1999) Policies to foster human capital. NBER working paper 7288. National Bureau of Economic Research, Cambridge, Massachusetts.

Heckman J, Klenow P (1998) Human capital policy. In: Boskin M (ed) Policies to promote capital formation. Hoover Institute, Washington, DC.

Im F, Rosenblatt D (2013) Middle-income traps: A conceptual and empirical survey. Policy research working paper 6594. World Bank, Washington, DC.

Jones C (2002) Sources of U.S. growth in a world of ideas. American Economic Review 92(1):220–239.

Jones C, Vollrath D (2013) Introduction to Economic Growth W.W. Norton, New York.

Kraay A, McKenzie D (2014) Do poverty traps exist? Assessing the evidence. Journal of economic perspectives 28(3):127–145.

Krueger A (2002) Economic considerations and class size. NBER working paper 8875. National Bureau of Economic Research, Cambridge, Massachusetts.

Krueger A, Lindahl M (2001) Education for growth: Why and for whom? Journal of economic literature 39(4):1101–1136.

La Ferrara E (2003) Model of credit transactions in Ghana. American economic review 93 (5):1730–1751.

La Porta R, Shleifer A (2014) Informality and development. Journal of economic perspectives 28 (3):109–126.

Lucas R (2002) Lecture on economic growth. Harvard University Press, Cambridge, Massachusetts.

McMillan M, Rodrik D (2011) Globalization, structural change, and productivity growth. In: Bacchetta M, Jansen M (eds) Making globalization sustainable. International Labour Organization and World Trade Organization, Geneva.

Miguel E, Kremer, M (2004) Worms: Identifying impacts on education and health in the presence of treatment externalities. Econometrica 72(1):159–217.

Oster E, Shoulson I, Dorsey E (2013) Limited life expectancy, human capital, and health investments. American economic review103(5):1977–2002.

Quataert D (1993) Ottoman manufacturing in the age of the industrial revolution. Cambridge University Press, Cambridge, UK.

Rangazas P (2002) The quantity and quality of schooling and U.S. labor productivity growth (1870–2000). Review of economic dynamics 5(4):932–964.

Rangazas P (2005) Human capital growth: An alternative accounting. Journal of macroeconomics 5 (1):1–43.

Ranis G, Fei J (1961) A theory of economic development. American economic review 51 (4):533–565.

Rodrik D (2011) The future of economic convergence. NBER working paper 17400. National Bureau of Economic Research, Cambridge, Massachusetts.

Rodrik D (2013) Structural change, fundamentals, and growth: An overview. Institute for Advanced Studies, Princeton, New Jersey.

Rodrik, D., 2015, Premature Deindustrialization, Working Paper, Harvard University.

Rogerson, R., 2008, Structural Transformation and the Deterioration of European Labor Market Outcomes, *Journal of Political Economy*, 116, pp. 235–259.
Ruggles S (2001) Living arrangements and the well-being of older persons in the past population. Bulletin of the United Nations 42/43, pp 111–161.
Ruggles S (2005) Intergenerational coresidence and economic opportunity of the younger generation in the United States, 1850–2000. Minnesota Population Center Working Paper.
Shavit Y, Pierce J (1991) Sibling size and educational attainment in nuclear and extended families: Arabs and Jews in Israel. American sociological review 56(3):321–330.
Strauss J, Thomas D (1998) Health, nutrition, and development. Journal of economic literature 34 (2):766–817.
Vijg J (2011) The American technological challenge. Algora Press, New York.
Wantchekon L, Novta N, Klasnja M (2013) Education and human capital externalities: Evidence from colonial Benin. Department of Politics, Princeton University.

10 Urbanization

In this chapter we study migration to the city and its effects on *urbanization*. In previous chapters we studied how the structural transformation affects economic growth and, in particular, how migration to the modern sector may alter private sector behavior. Here, we focus on the question of the best pace of urbanization as it relates to the allocation of rural and urban government services. Our motivation comes from the fact that the vast majority of governments around the developing world are concerned about the adequacy of public goods provision and the crowding associated with rapid urbanization (Bloom and Khanna (2007)).[1] In this sense, the structural transformation, which generally raises economic growth, can occur too quickly. A second important issue we address is the role politics plays in exacerbating rural-urban inequalities. As first stressed by Lipton (1977), the disproportionate political power of urban interests (the "urban elite") in some developing countries' economic policies may distort the allocation of government services, exacerbate rural-urban inequalities, and intensify migration beyond efficient levels.[2]

To investigate the issues surrounding regional migration and the allocation of government services we use our dual economy framework by interpreting the traditional sector as rural and the modern sector as urban. The rural technology is traditional in the sense that land and labor are used as inputs. The urban technology is modern in the sense that physical capital and labor are used as inputs. Thus, similar

[1]The concern is so significant that 73% of the governments surveyed have designed policies with the goal of slowing urbanization. A major concern over urbanization is the ability to finance the necessary public infrastructure associated with a growing urban population. According to some estimates, trillions of dollars of investment will be needed to extend the urban infrastructure due to the high urbanization rates in Asia.

[2]In addition to providing a voting-based theory of urban bias, Majumdar et al. (2004) empirically document the extent to which public services in urban areas exceed those in rural areas. McCormick and Wahba (2003) examine the effects of an urban bias in the allocation of public sector jobs. Bezemer and Headey (2008) discuss continuing concerns over the presence of an urban bias in domestic and international development policy. Bates (1981) focuses on how urban bias affects the pricing of agriculture products.

S. Das et al., *Economic Growth and Development*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-319-89755-4_10

to Chap. 9, the rural and urban sectors are distinguished by the different technologies used to produce a common good.[3]

The main contribution of this chapter is to introduce fiscal policy into the dual economy framework. To abstract from the effects of the structural transformation on private capital intensity, a major focus of Chaps. 8 and 9, we study a government operating in a small open economy. We also simplify the analysis by abstracting from explicitly modeling human capital and endogenous fertility in order to concentrate on endogenous fiscal policy. The policy maker's problem is how to allocate a given budget for productive services across the rural and urban sectors. The government inputs are interpreted as any service that raises productivity either through human capital development (e.g. training, education, or public health programs), physical infrastructure (e.g. roads, bridges, or public utilities), or property protection (e.g. policing or fire protection services). As is commonly alleged, we assume that the policy maker "favors" urban households over rural households.

In our setting, if the government also has the ability to directly allocate labor across sectors then the first-best, productively efficient, outcome would require no urban bias in the allocation of productive government services. If both sectors operate, an equal allocation of productive services is first-best, equating the marginal productivity of both labor and government services across the two sectors. Thus, the requirement of an urban bias is tied to the fact that the government cannot directly allocate labor across sectors when maximizing total output.

When households choose their location, they compare the relative wage opportunities in the rural and urban sectors. There is a wage gap in this chapter because it is costly to migrate from the traditional to the modern sector. If the wage gap across sectors is sufficiently high, then some households in the low-wage rural sector incur the cost of moving to the high-wage urban sector. We focus on the commonly observed historical pattern in developing countries where low-wage rural households migrate to the high wage urban sector over long periods of time.

In the presence of endogenous rural-urban migration based on household choice, we find that the second-best allocation of government services across rural and urban sectors is independent of the weight the policy maker places on rural households; politics plays no role in the allocation. Nevertheless, we find that the second-best efficient allocation of government services is "biased" toward the urban sector.

The efficient urban bias—more government services per capita are available in the city than in the country—is driven by two key features of the economy. First, because of unrestricted migration across sectors, there is a close link between rural

[3]For the purpose of the issue that we address, the apparent policy bias against the poorer and less developed rural sector during the development process, distinguishing the sectors by differences in traditional and modern technologies seems reasonable. There are certainly other differences between the rural and urban sectors that we ignore. Also by focusing on an "urban sector" we do not address the expansion of cities across geographic areas or heterogeneity in city types. For an analysis of systems of cities, where cities specialize in the production of different goods, see Henderson (1974), Black and Henderson (1999), and Duranton and Puga (2004), and Henderson and Wang (2005).

and urban wages. An increase in the quantity of government services in the city raises the productivity and wages of urban workers. A higher urban wage increases rural-to-urban migration and thereby raises rural wages as well. Second, migration to the city does not directly alter the urban wage. Migration to the city, where the modern technology is operated, raises the marginal product of private capital. In an open economy, the higher marginal product of capital causes an inflow of foreign capital that maintains the interest rate and the urban wage rate. Therefore, migration to the city only affects wages through a crowding of urban government services per capita. Taken together, these two features suggest that to maximize wages across the economy, the government must choose the allocation of productive services to generate a migration flow that maximizes *urban* government services per capita.

The fact that rural government services have an effect on migration suggests that it is not efficient—even for a government dominated by urban households—to spend the entire government budget on urban services. Such a policy would intensify migration, placing a strain on the government budget that reduces government service per capita in the city and wages in both sectors. Instead, the efficient policy—the one that maximizes urban government service provision per capita and wages in both sectors—is to provide some government services in the rural area in order to limit migration.

Given the urban bias does not depend on the weight the government places on rural household welfare, then how is redistribution effected if, as Lipton argued, politics of developing countries are often dominated by urban elites? In our model, a "redistributive" urban bias designed to increase urban welfare must take the form of restricting migration to the city. To redistribute income, the government needs to reduce migration by imposing direct administrative measures, such as migration quotas and other restrictions that raise the cost of migration.[4] We show that an increase in migration restrictions, even allowing for a policy response in the allocation of government services, will raise urban wages and lower rural wages. This result suggests that political debates are more likely to center on migration restrictions rather than on the allocation of government goods.

Section 10.1 presents a simple baseline model that generates an unambiguous urban bias in government service provision on efficiency grounds. Section 10.2 analyzes the effects of economic growth, stemming from both balanced and unbalanced technological change, on migration and the rural-urban wage gap. Section 10.3 discusses several extensions of the basic model. Section 10.4 discusses the recurrent problem of slums that form in the cities of developing economies. Section 10.5 ends with some concluding remarks.

[4]Clear examples of where such policies have been carried out include the former Soviet Union and China. See also Sect. 10.6

10.1 Urban Bias

Our baseline economy is small and perfectly open to both the trade of goods and the flow of physical capital across borders.[5] There are two sectors of production that produce the same good, urban and rural, of fixed geographic size. Production in each sector is carried out by perfectly competitive firms. Households can only work in the sector in which they reside, but they choose to locate in the rural or urban sector. Locational choices are made by young households before work begins. We focus on the typical pattern seen over the course of development, where young households from the rural sector migrate to the urban sector over relatively long periods of time in response to wage gaps between sectors.

The fundamental difference between sectors is the technology used in production. The technology used in the rural sector is traditional in that land is combined with labor to produce output, while a modern technology is used in the urban sector where physical capital and labor are combined to produce output. It is important to note that because the same good is produced in each sector, one sector may be viewed as being redundant. In particular the policy makers favoring the urban sector may choose to ignore the rural sector altogether by failing to provide government services there. Feler and Henderson (2011) discuss how some governments enact regulations that withhold public services from low-income households to slow migration to the city.

10.1.1 Production

10.1.1.1 Urban Sector

The urban sector technology for producing goods is given by

$$Y_t = K_t^{\alpha}(D_t M_t)^{1-\alpha} \tag{10.1}$$

where Y is output, K is the capital stock, M is the number of workers employed, D is a labor productivity index, and α is the constant capital share parameter. Similar to Chap. 3, the productivity index is a function of effective productive government services per worker, G_t/M_t, and an exogenous technology index, E,

$$D_t = \left(\frac{G_t}{M_t}\right)^{\mu} E_t^{1-\mu} \tag{10.2}$$

where μ is a constant parameter. The technology index, E, grows at a constant rate q.

Profit maximization and competition generate the standard equations relating factor prices to the marginal productivity of the factors

[5] The next three sections are based on Mourmouras and Rangazas (2013). In an unpublished appendix, we consider the closed economy case. The appendix can be found as a supplement to the article on the *Journal of Economic Geography's* website.

$$r = (1-\tau)\alpha k_t^{\alpha-1} \tag{10.3a}$$

$$W_t = (1-\alpha)D_t k_t^{\alpha}, \tag{10.3b}$$

where $k_t \equiv K_t/D_tM_t$. The return to capital is taxed at the constant rate τ. Capital depreciates at the constant rate δ. For simplicity, we take the rate of depreciation to be one. The internationally determined after-tax rate of return on capital is r. Wage income is also taxed at the constant rate τ, but the tax is collected at the household level. The before-tax wage rate paid to a worker is W which, unlike w from previous chapters, incorporates the return to a unit of human capital and the effect of the labor-augmenting productivity index.

10.1.1.2 Rural Sector

The rural sector produces the same good (denoted by O only for the purpose of helping to keep the two sectors distinct) with a technology that uses land (L) instead of physical capital. For simplicity, we initially assume that rural firms and households face the same tax rate as in the urban sector. Later we consider the more realistic situation, initially discussed in Chap. 8, where taxes are more difficult to collect from the rural sector than they are from the urban sector.

The rural technology is

$$O_t = L_t^{\alpha}\left(\tilde{D}_t F_t\right)^{1-\alpha} \tag{10.4}$$

where F is the labor employed in production and $\tilde{D}$ is the labor productivity index in the rural sector. In keeping with evidence that the labor share of income does not vary with the composition of output over the course of development, we set the land share to be the same as the capital share in the urban sector, α. Similar to the urban sector, the labor productivity index in the traditional sector is determined by

$$\tilde{D}_t = \left(\frac{\tilde{G}_t}{F_t}\right)^{\mu}\left(\tilde{E}_t\right)^{1-\mu} \tag{10.5}$$

where $\tilde{E}_t$ is the technology index in rural production.[6]

The competitive factor price equations for the rental rate of land and the rural wage rate are

$$r_t^L = (1-\tau)\alpha\left(L_t/\tilde{D}_t F_t\right)^{\alpha-1} \tag{10.6a}$$

$$\tilde{W}_t = (1-\alpha)\tilde{D}_t\left(L_t/\tilde{D}_t F_t\right)^{\alpha}, \tag{10.6b}$$

[6]Note that we are modeling government services as publically provided private inputs. In some cases it is more appropriate to treat government services as flowing from impure public goods. We examine this case in Sect. 8.3.

Available land is fixed in the economy, so in equilibrium we must have $L_t = L$ in each period. Given the equilibrium rural wage rate and the fixed amount of land, the allocation of labor to the rural sector, F_t, and the rental rate on land, r_t^L, are given by the competitive factor price equations, (10.6a, 10.6b).

10.1.2 Households

Households are life-cycle planners that populate the standard overlapping generations model. Households live for two periods; they work only in the first period and then retire in the second period. All households have the same preferences. Household welfare is determined by the consumption of the common good produced in the two sectors, c_{it}, where i denotes the period of life in which the goods are consumed. Households save by purchasing physical capital (s_t) and land (l_t), and then renting them to producers. Households can only work in the sector in which they reside. Unlike in Chaps. 7, 8, and 9, there is no constraint on hours in the rural sector and land is traded in perfectly competitive markets.

10.1.2.1 Urban Sector

Household preferences, regardless of location, are given by the following lifetime utility function

$$U_t = u(c_{1t}) + \beta u(c_{2t+1}), \tag{10.7}$$

where u is a strictly concave and differentiable, single-period utility function, and β is the time discount factor.

The single-period budget constraints for the two periods of life of an urban household are

$$c_{1t} + s_t + p_t^L l_t = (1 - \tau) W_t \tag{10.8a}$$

$$c_{2t+1} = R s_t + \left(r_{t+1}^L + p_{t+1}^L\right) l_t, \tag{10.8b}$$

where p_t^L is the competitive relative price of land and $R \equiv 1 + r - \delta$. The two single-period constraints generate the following lifetime budget constraint,

$$c_{1t} + \frac{c_{2t+1}}{R} = (1 - \tau) W_t, \tag{10.9}$$

with the no-arbitrage condition that determines the equilibrium price of land,

$$p_t^L = \frac{r_{t+1}^L + p_{t+1}^L}{R}. \tag{10.10}$$

Given the competitive factor prices, urban households make life-cycle consumption choices to maximize utility.

10.1.2.2 Rural Sector

There are two types of rural households. One type chooses to remain in the rural sector and the second type migrates to the urban sector. Migration is costly. It requires a loss of consumption equal to ω_t (transportation costs, goods left behind. a drop in housing quality, and other moving expenses) and a lost fraction of work-time, $\overline{\omega}$, spent in transit and looking for work in city.[7] The decision to migrate occurs at the beginning of a household's working life and is based on the relative wage opportunities in the two sectors that determine the household's lifetime welfare.

Households that remain in the rural sector maximize utility subject to a lifetime budget constraint based on the after-tax rural wage,

$$\tilde{c}_{1t} + \frac{\tilde{c}_{2t+1}}{R} = (1 - \tau)\tilde{W}_t.$$

The second type of rural household migrates to the urban sector. Migrants must bear the time and goods costs associated with migration. They maximize utility subject to a lifetime budget constraint based on the urban wage net of migration costs,

$$\bar{c}_{1t} + \frac{\bar{c}_{2t+1}}{R} = (1 - \tau)\left(1 - \overline{\omega}\right)W_t - \omega_t.$$

To assess household welfare in each sector we use a value function giving the maximum household lifetime utility as a function of the market factor prices. In an open economy, the interest rate is determined exogenously by the international market for capital. We assume the international return to capital is constant, so that maximum utility will vary because of variation in the market wage rate alone. The value function is defined to be $V(X)$, where X stands in for the net wage of the different household types. In some cases we will resort to the special case of log preferences that we employed heavily in past chapaters, where $u(c) = \ln c$. In this case the value function of a household with net wage X takes the form,

$$V(X) = (1 + \beta)\ln\left(\frac{1}{1 + \beta}\right) + \beta\ln\beta + (1 + \beta)\ln(X) + \beta\ln(1 + r).$$

10.1.3 Demographics

Each period there will be young and old households of each of the three types. The population of young urban households in each period, N_t^*, is comprised of the children of last period's urban-sector natives and the young rural households who choose to migrate

[7]The lost work-time associated with migration can also be given a Harris and Todaro (1970) interpretation in that those arriving in the city endure a period of search unemployment.

$$N_t^* = nN_{t-1}^* + (1 - \varphi_t)n\tilde{N}_{t-1} \tag{10.11}$$

where $n-1$ is the common and exogenous rate of population growth for all households, φ_t is the fraction of young rural households that choose to remain in the traditional sector, and $\tilde{N}_{t-1}$ is the number of rural households last period. The number of rural households in the current period is given by

$$\tilde{N}_t = \varphi_t n\tilde{N}_{t-1}. \tag{10.12}$$

The labor supply in each sector equals the number of young households in that sector, $M_t = N_t^*$ and $F_t = \tilde{N}_t$. Assuming a common population growth rate in each sector implies the country's total population size $N_t \equiv N_t^* + \tilde{N}_t$ is exogenous, i.e. independent of the endogenous allocation of the population across the sectors that occurs each period.

10.1.4 Migration in Equilibrium

We primarily consider equilibria where both sectors operate and some movement to the city occurs each period. For these equilibria, the rural households must be indifferent about staying in the rural sector or migrating to the urban sector. To be indifferent about migrating, the value functions must be equal whether the household migrates or not, i.e.
$V_t\left((1-\tau)\tilde{W}_t\right) = V_t\left((1-\tau)\left(1-\overline{\omega}\right)W_t - \omega_t\right)$ or equivalently

$$\tilde{W}_t = \left(1 - \overline{\omega}\right)W_t - \frac{\omega_t}{1-\tau}. \tag{10.13}$$

Positive migration to the city is needed to satisfy (10.13) when, in the absence of migration, is strictly less than $\left(1 - \overline{\omega}\right)W_t - \dfrac{\omega_t}{1-\tau}$.

10.1.5 Government

To examine the possibility of an urban-bias in setting policy we allow urban households to determine the allocation of public services across the two sectors of the economy. Given that all urban households in a given age cohort are identical, the choice of government service allocation can be made by a representative household from the cohort of young households in each period. Only young households care about the allocation of government services because these services affect welfare only by affecting the productivity of labor and wages. The welfare of old households in each period is predetermined by the previous period's wage and the exogenously determined world interest rate.

The government budget constraint confronting the representative urban household in making its fiscal choice is given by

$$G_t + \tilde{G}_t = B_t, \tag{10.14}$$

where B_t is the portion of the total government budget that is allocated to fund government services. In this section we assume that the value of B_t is exogenous to the model and focus only on the allocation problem. We assume that B_t rises proportionately with the state of technology and the country's population, to capture the effects of an increasing tax base. In Sect. 8.3 we make tax revenue and the budget endogenous.

The representative urban household chooses the allocation of public services to maximize a social welfare function of the form $V_t((1-\tau)W_t) + \upsilon V_t\big((1-\tau)\tilde{W}_t\big)$, subject to (10.14), where υ is a nonnegative weight the policymaker places on the welfare of the rural household.

To focus on the consequences of an urban bias in politics we assume$0 \leq \upsilon \leq 1$. The government chooses the allocation of public services taking account the full general equilibrium interactions; in particular the determinants of the urban wage (10.3b), the link between urban and rural wages given by (10.13), and the effect of the government service allocation on the migration flow to the city, which is indirectly determined by (10.6b).

10.1.6 Efficient Urban Bias

The solution to the government's problem is

$$\tilde{g}_t = \frac{\mu(1-\alpha)}{\alpha + \mu(1-\alpha)} g_t, \tag{10.15}$$

where we define the de-trended, for exogenous technological progress, per worker values of $\tilde{G}_t$, G_t, and B_t as lower case values, that is government services per worker in each sector and the budget available per household, all de-trended by the state of technology.[8] Equation (10.15) determines the optimal mix of government services across sectors. Note that the expression is independent of υ, so politics plays no role in determining the allocation of government services across sectors. There is an unambiguous "urban bias" because government services per capita are smaller in the rural sector, but this is for efficiency reasons.

The fundamental logic for the efficient urban bias starts with the idea that urban wages are solely a function of urban public services per capita, see (10.3b), and not the absolute value of labor as in the rural sector where land is an input. This is true because the private capital-labor ratio is independent of migration flows to the city.

[8]For proofs of this result and others in this chapter see Mourmouras and Rangazas (2013).

As the city becomes more populated, the marginal product of capital rises and attracts private capital into the country to maintain interest rates at the world level. This causes the capital-labor ratio to remain constant and thus there is no crowding of private capital that would result in lower wages. Also note that the urban wage determines the rural wage when there is migration across sectors (see 10.13). These two features imply that maximizing government services per capita in the city maximizes wages in both locations.

This logic seems to suggest that rural public services should be set to zero. However, to maximize per capita public services in the city, migration must be limited to some extent by offering public services in the rural sector as well. Thus, the sole purpose of rural public service provision is to control urban crowding leading to the maximization of urban public services for a given fiscal budget.

The role of rural public services in controlling urban crowding can be seen directly by rewriting (10.14) as $g_t\pi_t + \tilde{g}_t(1-\pi_t) = b_t$. Given the size of the government budget per person, an increase in the fraction of the urban population (π_t) will strain resources when there is an urban bias. The value of π_t is indirectly determined by (10.6b), that expresses the demand for labor in the rural sector as a fraction of the rural population (φ_t) that does not leave for the city (see *Problem* 2). The lower is the provision of government services in the rural sector, the weaker is the demand for rural workers and the higher is π_t. Thus it is optimal to provide some government services to the rural sector in order to manage the value of π_t in face of an urban bias in government service provision.

The degree of urban bias depends on how much rural government services affect the rural population; the larger the effect, the weaker is the urban bias. The effectiveness of government services in controlling migration depends on the production parameters, μ and α. The larger is the expression $\mu(1-\alpha)$, the greater is the effect of government services on the marginal product of labor, and labor demand, in the rural sector. A low value of α not only increases the size of the "shift" in the demand for labor, as government services increase, but also reduces the slope of the demand for labor. When the demand for labor has a flatter slope, more workers must be hired in order to drive the marginal product of labor back down to the equilibrium wage rate.

The relative wage inequality across sectors is determined by (10.13). The amount of wage inequality that rural households will tolerate depends on the costs of migration relative to the benefit of migration, the urban wage rate. Wage inequality will decrease over time if the urban wage rate increases because it lowers the relative "goods" costs of migration (transportation costs, land or goods left behind, difference in housing costs across sectors, etc.).

Finally, note that the efficient urban bias result depends on the assumption that there is a positive migration flow across sectors. If migration was impossible, or perfectly restricted, then the policy maker's preferences *would* affect the allocation of public services. For example, in the extreme case where the policy maker was only maximizing the preferences of urban households, the entire government budget would be devoted to urban public services because, with migration restrictions, the crowding effects would no longer be an issue.

More generally, with no migration, υ will influence the size of the bias. For example, if consumers have log preferences the policy solution is $\tilde{g}_t = \frac{\upsilon \pi_t}{1 - \pi_t} g_t$, where the extent of the bias depends on the relative cost of providing the services, which are a function of the relative sizes of the urban and rural populations, and on υ.

10.2 Growth and Urbanization

In this section we consider how various events affect urbanization and growth in the model developed in Sect. 10.1.

10.2.1 Urbanization with Balanced Growth

The economy's dynamics is determined by the exogenous pace of technological progress (E_t) and the extent to which the progress is balanced across sectors, i.e. the extent to which $\tilde{E}_t$ keeps pace with E_t. Balanced technological progress creates a "pull-factor" that increases urbanization but lowers the intensity of productive government services economy-wide. The direct effect of an increase in E_t raises wages, lowers the relative cost of migrating, and reduces wage inequality (see 10.13). However, because an increase in E_t also lowers g_t, the overall effect on inequality is not obvious. One can show that the indirect effect of a lower g_t, only mediates, rather than completely offsets, the direct effect of a higher g_t.

Urbanization caused by balanced technological progress supports the "optimistic" view of urbanization.[9] In this case, urbanization is a natural consequence of progress and increasing opportunity. Some crowding of services results, but the level of wages increase and wage inequality falls. Here, urbanization is associated with rising worker productivity.

As E_t increases over time the fraction of the rural population that leaves for the city each period also increases, accelerating the growth in the fraction of the entire population that works and lives in the city. Along the growth path, the urban-rural wage gap becomes smaller. Eventually, the traditional rural sector disappears, as the entire economy converges to the modern urban sector (as in Hansen and Prescott 2002; Lucas 2004).

It is important to stress that an increase in technology in the urban sector *only* would not necessarily deliver this favorable outcome. If E_t increases and $\tilde{E}_t$ does not keep pace, a larger migration to the city results. The larger migration may lower g_t enough to offset the direct effect of E_t on urban wages. Balanced technological progress is required to guarantee that urbanization is associated with rising worker productivity and falling wage inequality.

[9] See Bloom and Khanna (2007) for a discussion of the optimistic and pessimistic views of urbanization.

10.2.2 Urbanization Without Balanced Growth

As documented by Fay and Opal (2000) there is a "pessimistic" view of urbanization, one revealed by the concerns of politicians who attempt to limit migration to the city (see the introduction). The pessimistic view is supported by the model when the underlying reasons for urbanization are ones that lower the relative productivity in the rural areas and push people toward the city—for example a rise in the rural population or relatively slow technological progress in the rural sector. These factors result in crowding and lowerg_t, for any given value of E_t. The reduction in government service intensity lowers wages in both sectors and increases wage inequality. Here, urbanization is associated with falling worker productivity.

The intuition about unbalanced technological change that favors the urban sector extends to physical capital inflows. Physical capital favors the urban sector by raising productivity in the city relative to the rural sector. For this reason, relative to its direct impact on urban wages, the effect of physical capital on encouraging migration to the city is large. Thus, the negative effect of reducing government services per worker may be large enough to offset the direct effect on productivity, and urban wages may fall.

When urbanization occurs without growth, there are policies that slow migration and improve the welfare of all households. An example of such a policy would be to direct foreign aid toward the rural sector that is designed to improve the relative state of rural technology by raising $\tilde{E}_t$. Such a policy would raise φ_t and g_t. This would result in an increase in government services in both sectors, higher wages in both sectors, and less wage inequality. Thus, directing aid projects to the rural sector will benefit households in both sectors.

It is important to note that, without a feedback of the urban population share on fiscal policy, increasing the level of technology in the rural sector would *not* raise welfare. If g_t were fixed, then improving rural technology would create a Malthusian-type increase in the population of the rural sector that would completely eliminate any productivity gains for an individual worker. It is the fact that lowering the urban population share raises effective government services in the urban sector that reduces wage inequality. The rise in urban government services increases urban wages and limits the population expansion in the rural area, allowing the rural wage to rise as well.

Figures 10.1 and 10.2 give an example of the country's dynamics with and without balanced growth. We consider two economies each with annual population growth rates of 2%. The population growth lowers rural wages and "pushes" the population toward the urban sector. In one economy there is balanced technological change leading to 1% growth in real wages, represented in the figures by the solid line, and in the other economy there is no technological change, represented by the dashed line. Each period in the model represents 30 years. In the initial period, 20% of the population is in the urban sector and the rural wage is 80% of the urban wage. We set $\alpha = \mu = 1/3$, so by (10.15) productive expenditures in the rural sector are 40% of urban expenditures throughout the economy's development.

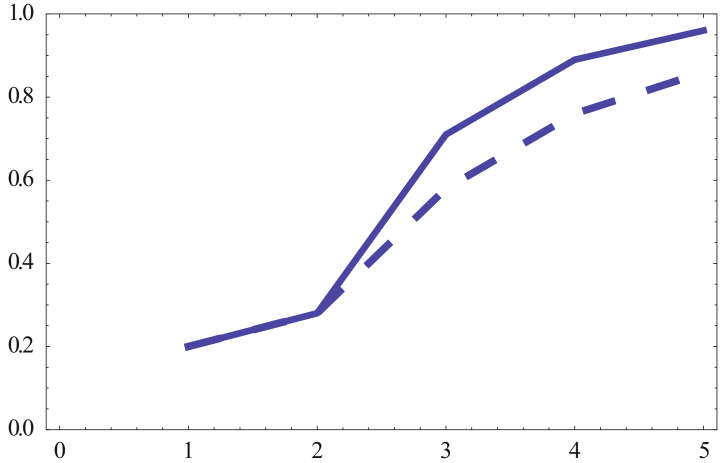


Fig. 10.1 Fraction of population in urban sector
Notes: Solid line is for economy with 1% annual growth in *E* and dashed line is for economy with zero growth

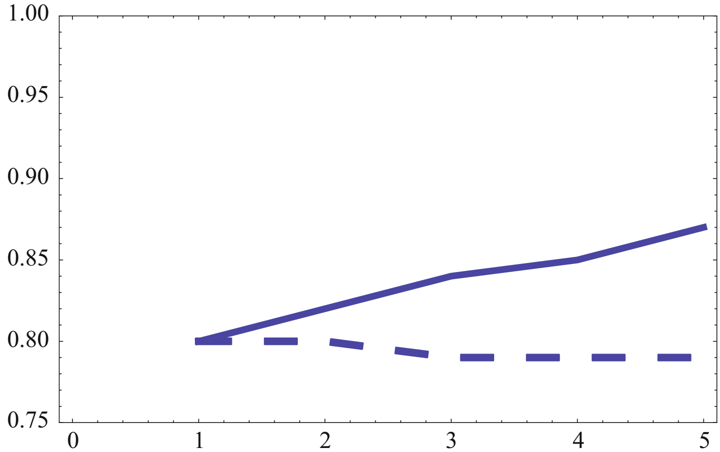


Fig. 10.2 Wage gap—short-fall of rural wages
Notes: Solid line is for economy with 1% annual growth in *E* and dashed line is for economy with zero growth

Without balanced growth, households are "pushed" into the urban sector by population growth and the wage gap worsens slightly over time. With balanced growth, a "pull" factor is added that increases the rate of urbanization relative to the no growth case. After 120 years the fraction of the population in the urban sector is 96% compared to 86% without balanced growth. More importantly, even with modest growth, the rural sector wage gap narrows significantly to 87% of the urban wage after four periods (120 years).

10.2.3 Redistributive Urban Bias

It is not in the interest of the policymaker to deviate from the efficient urban bias given by (10.15), even when the weight placed on the welfare of the rural household is zero. However, the government may add to the natural costs of migration by artificially increasing the cost. This is where one would expect to see political debates centered. An increase in ω_t raises g_t and increases productivity and wages in the urban sector. Despite the fact that $\tilde{g}_t$ rises as well, productivity and wages in the rural sectors fall, increasing wage inequality.

Thus, increasing the cost of migration, whether directly due to government policy or not, will lessen urban crowding and give rise to an increase in public services across the economy. However, in the rural sector, the effect of rising public services will not be enough to keep wages from falling as fewer workers choose to migrate because of the higher cost resulting in the crowding of land and lower worker productivity. Thus, the politics of disproportional urban influence in the setting of policy will take the form of raising artificial restrictions on migration or by not taking actions that reduce the natural costs of migration. While migration restrictions are not efficient, because they trade-off the welfare of rural and urban households, they could increase *aggregate* welfare. Rangazas and Wang (2018) provide examples where positive migration restrictions raise aggregate welfare, even when the utility of all households are weighted equally.

10.3 Extensions

10.3.1 Government Transfers

Our focus has been on productive government spending largely because of the concerns about the crowding of urban infrastructure that were discussed in the introduction. However, it is natural to ask how the presence of government consumption and transfers may affect the results. We first consider the case where government consumption is a substitute for private consumption (e.g. food and housing subsidies). This case is equivalent to the situation where the government directly provides cash transfers or wage subsidies to households, a common form of alleged favoritism in some developing countries, if the public provision of the private goods is not too large a fraction of the household budget. We will also consider a different situation where the government consumption is not a close substitute for private consumption and thus is very different from government cash transfers to households (e.g. vaccinations that create direct consumption benefits from improved health).

Suppose that the government provides a consumption-good to households in each period of life: Q_{1t} and Q_{2t+1} for an urban household of generation-t and $\tilde{Q}_{1t}$ and $\tilde{Q}_{2t+1}$ for a rural household of generation-t. In the case where the government consumption is a substitute for private consumption, total consumption of good i for the urban household is $\widehat{c}_{it} \equiv c_{it} + Q_{it}$, and similarly for the rural household. We will consider the case where the government good provision is not too large so that it is equivalent to a cash transfer or wage subsidy. The case where this is not true, i.e. when the government effectively chooses the consumption level of a good for the household, is similar to the next case we look at where the government consumption good is not a substitute for private consumption.

When the government consumption good is a substitute for private consumption, the urban household's behavior, and similarly for the rural household, can be modeled as maximizing

$$u\left(\widehat{c}_{1t}\right) + \beta u\left(\widehat{c}_{2t+1}\right)$$

subject to

$$\widehat{c}_{1t}\frac{+\widehat{c}_{2t+1}}{R} = (1-\tau)W_t + \left(Q_{1t} + \frac{Q_{2t+1}}{R}\right).$$

The timing of when the household receives the consumption good/transfers is irrelevant, so for notational simplicity we consider the case where the government provides the consumption goods in the first period of life only.

In this situation, the analog to Eq. (10.13), the condition needed to generate partial migration, is

$$(1-\tau)\tilde{W}_t + \tilde{Q}_{1t} = (1-\tau)\left(1-\overline{\omega}\right)W_t - \omega_t + Q_{1t}.$$

So, differences in government consumption/transfers across the two regions affect the migration choice.

The government's optimization problem, which now involves choosing the de-trended government consumption goods q_{1t} and $\tilde{q}_{1t}$, yields the following expression that informs us about urban bias.

$$\frac{q_{1t} + g_t}{\tilde{q}_{1t} + \tilde{g}_t} - 1 = \frac{\alpha}{\mu(1-\alpha)}\frac{\tilde{g}_t}{\tilde{q}_{1t} + \tilde{g}_t}. \tag{10.16}$$

The first thing to notice about (10.16) is that the basic message of the early analysis carries over to this setting; there must be an urban bias and the weight the government places on the rural household has no bearing on the presence of the bias and no direct effect on the size of the bias. The logic for this result is the same as before. The government should attempt to maximize the welfare of the urban household, which indirectly maximizes the welfare of the rural household when there is a positive migration flow. However, some provision to the rural sector is needed to control urban crowding of both productive consumption services and government consumption.

There are some nuances associated with (10.16) that were not present before. First, the necessary bias is based on *total* government spending in the urban sector relative to the rural sector. It is now possible that there is no bias in the provision of *productive* government inputs. Instead the government may favor the urban sector in terms of *consumption* good provision and transfer payments. Second, the *size* of the urban bias is inversely related to the *form* of the provision of goods to the rural sector. In particular, the greater is the relative amount of consumption goods provided to the rural sector, the smaller is the total urban bias. Through this avenue the size of the bias may vary with the weight the policy makers place on the rural households. So while the presence of a bias is independent of the weight, the extent of the bias may now vary with the weight.

To see this second point more explicitly, first recognize that under our assumptions some productive government service provision is needed to operate the rural sector. While this assumption is extreme it is meant to capture the idea that the return to public investment in production is high when the level of productive services is low—i.e. the return to the provision and maintenance of basic infrastructure such as elementary education and simple roads is high. This means that in raising the quality of life in the rural sector to avoid crowding in the urban sector, the government should start with these basic productive services. However, if the weight that the government places on the rural sector is relatively low, then the government may not go beyond that, i.e. it may choose not to provide consumption goods or transfers to the rural sector so that $\tilde{q}_{1t} = 0$. In this situation, the second expression on the right-hand-side of (10.16) is always one and we can solve (10.16) to get an expression that is very similar to (10.15),

$$\tilde{g}_t = \frac{\mu(1-\alpha)}{\alpha + \mu(1-\alpha)}(g_t + q_{1t}).$$

Here, the urban bias is strictly independent of υ, at least over some range of υ and perhaps over the entire range. In addition, the primary way that the government may be viewed as favoring the urban sector is in the provision of government consumption goods and transfers.

However, it is also possible that at a sufficiently high value of υ, but still below one, that the government will provide some consumption goods, so that $\tilde{q}_{1t} > 0$. For this range of υ, the second expression on the right-hand-side of (10.16) is always less than one and the size of the urban bias must fall relative to the case when $\tilde{q}_{1t} = 0$. So, while an urban bias must be present for any υ, it is now possible that the size of υ affects the size of the bias.

Next, suppose that the government consumption good is not a perfect substitute for private consumption. In this case (10.16) remains a necessary condition for the government's optimal policy. However, now the government consumption good may also be viewed as "essential" and thus must always be positive. In this case the size of the urban bias would vary over the full range of υ.

10.3.2 Endogenous Taxation

To this point we have analyzed fiscal policy as a pure allocation problem, with an exogenous budget for total government service expenditures. We now show that the presumption of an efficient urban bias becomes stronger if one allows for endogenous tax revenue. For simplicity, we return to the situation where the government is choosing only productive services.

Instead of an exogenous budget we now assume that the budget for productive government spending is a fixed fraction $\overline{\eta}$ of total taxes collected, $\tau\left(Y_t + \overline{\sigma}O_t\right)$. Note that the tax base is endogenously determined because it depends on production from the two sectors and thus on the allocation of labor and productive government services across sectors. In addition, the parameter $\overline{\sigma}$ is bounded between zero and one, taking into account that collecting taxes from the traditional rural sector's tax base may be more difficult than in the modern urban sector.

With a Cobb-Douglas technology, we can write government revenue as

$$\tau\left(\frac{W_t M_t}{1-\alpha} + \frac{\overline{\sigma}\tilde{W}_t F_t}{1-\alpha}\right)$$

As before, wages in both sectors are independent of direct influence from $\tilde{g}_t$. However, now $\tilde{g}_t$ affects the tax base through its influence on φ_t and the relative size of the workforce across rural and urban sectors.

Solving the government's problem gives us the analog to (10.15) when the tax base is endogenous.

$$\tilde{g}_t = \frac{\mu(1-\alpha)}{\alpha + \mu(1-\alpha)}\left[g_t - \frac{\overline{\eta}\tau}{1-\alpha}\left(\frac{W_t - \overline{\sigma}\tilde{W}_t}{E_t}\right)\right] \tag{10.15$'$}$$

Comparing (10.15′) to (10.15), one can see that, with a wage gap in favor of the urban sector, the efficient urban bias is now stronger (i.e. $\tilde{g}_t$ is now a smaller fraction of g_t). The stronger urban bias results from the fact that an increase in $\tilde{g}_t$ carries an additional cost—a smaller tax base and a loss in tax revenue. An increase in $\tilde{g}_t$ reduces φ_t and increases the relative size of the rural sector. With a wage gap in favor of the urban sector, an increase in the relative size of the rural sector causes the tax base to shrink.

The argument just made would hold even if $\overline{\sigma} = 1$. Allowing for differential ability to collect taxes across sectors further increases the efficient urban bias. In developing countries, it is relatively difficult to tax the traditional sector. The effective tax rate in that sector is less than in the modern urban sector. Assuming that there remains a gap in the pre-tax wages in favor of the urban sector, this adds an additional reason why tax revenue would fall with an increase in $\tilde{g}_t$.

One can go further by substituting (10.15′) back into the government budget constraint to get a complete solution for spending in the two sectors as a function of tax revenue. In particular, the solution for $\tilde{g}_t$ is

$$\tilde{g}_t = \frac{\alpha}{\alpha + \mu(1-\alpha)} \frac{\mu(1-\alpha)}{\pi_t \alpha + \mu(1-\alpha)} \left[\frac{\overline{\eta}\overline{\sigma}\tau}{1-\alpha} \left(\frac{(1-\pi_t)\tilde{W}_t}{E_t} \right) \right]$$

Note that the term outside the square brackets is less than one because $0 \leq \pi_t \leq 1$. The term in the square brackets gives the tax revenue collected from the rural sector (on a per worker basis and de-trended for growth, as is the case for $\tilde{g}_t$). This means it is optimal for rural government spending per worker to be less than rural taxes collected, i.e. the rural sector must cross-subsidize urban government spending.

To justify rural spending, another dollar spent on the rural sector must create a net revenue gain of more than a dollar by reducing the urban population. This means the urban-rural spending gap must be *greater* than the urban-rural tax collection gap so that, as the urban population falls with an increase in rural spending, a marginal gain in revenue is generated. In other words, some of the rural sector taxes must be used to finance urban spending for there to be a net marginal gain in revenue to the urban sector from an increase in rural sector spending. Since the objective is to maximize urban wages via urban government spending, the rural sector must subsidize the urban sector. In fact if the rural sector cannot provide net revenue to the government, i.e.$\overline{\sigma} = 0$, then the government should ignore the rural sector, $\tilde{g}_t = 0$ and thereby force everyone into the city.

10.3.3 Impure Public goods

We have assumed that government services are private goods, which is appropriate for some but not all government services. Here we consider the case where government services flow from impure public goods.

Suppose now that effective productive government services are generated from impure public goods that are subject to crowding and thus are imperfectly shared by the workforce. As before we define G_t to be a measure of productive government services in the urban sector but now effective services are defined as $g_t = G_t/M_t^{\xi}$, where ξ is a constant parameter that lies between zero and one. The value of ξ determines the degree of crowding of public goods. If $\xi = 0$, then G_t is generated from pure public goods that are not subject to crowding. If $\xi = 1$, then G_t is generated from pure private goods, and to maintaing_t, G_t must rise proportionally with the workforce. Here we focus on the intermediate case where ξ is between zero and one, i.e. where G_t is generated from impure public goods that are subject to some crowding.

In similar fashion, we define effective government services in the rural sector to be $\tilde{g}_t = \tilde{G}_t/F_t^{\varepsilon}$. Due to its smaller geographic area, the urban sector is assumed to have an advantage in the sharing of the services generated by a given impure public good. The "public" nature of a good is partly a function of its characteristics and partly a function of the population density located around the good. For (i) a given local government public good provided to each sector and (ii) a given population of workers in each sector, the urban sector provides at least as many effective services per worker, i.e. we assume $0 \leq \xi < \varepsilon \leq 1$. For example, suppose the *same* size road is provided to the urban area and the rural area, and each area contains the *same* total population. It is more likely that a larger fraction of the urban working population lives close enough to the road to use it productively. Thus, there is an efficiency advantage to urbanization that directly stems from the provision of a shared public good.[10] Note that it remains true that an increase in the population in each sector will reduce government services per worker.[11]

Under these assumptions one can solve the government's problem to get the following equation for efficient allocation of government services

$$\tilde{G}_t/F_t = \frac{\xi\mu(1-\alpha)}{\alpha + \varepsilon\mu(1-\alpha)}(G_t/M_t). \qquad (10.15'')$$

Equation (10.15″) gives the implied relationship between the *observed* government services per worker in the two sectors (found by dividing a measure of the total government services by the population of workers). Recalling that $\xi \leq \varepsilon$, the equation indicates that there will always be an observed urban bias in the setting of fiscal policy, $G_t/M_t > \tilde{G}_t/F_t$. In addition to the other features identified under the assumption that government services are private goods, the degree of urban bias also now depends on how advantageous it is to share the impure public goods in the city versus the country—i.e. the lower the value of ξ/ε the larger is the bias.

More important than the observed bias is the *effective* bias per worker, $\tilde{g}_t/g_t$, a determinant of the relative TFP in the two sectors. The effective bias is given by

[10] Cities may also contain other efficiency advantages such as improved matching between employee and employer and learning spillovers from increased worker interaction (Duranton (2008)). Rosenthal and Strange (2004) point out that identifying the precise source of these efficiency advantages is extremely difficult. Moretti (2004a) discusses the difficulty in establishing the presence and magnitude of human capital externalities in cities. Henderson (2010) argues that urbanization per se does not increase growth rates, although growth rates may increase when the population becomes concentrated in very large cities. In Chap. 7 we mentioned evidence suggesting that there are learning externalities related to the average human capital per person in the community where one lives. Moretti (2004b) and Liu (2014) find that the average human capital per worker in a city raises the productivity of firms in the city, especially those that employ high human capital workers. Duraton (2014) reviews the literature and concludes that cities have a positive causal effect on worker productivity.

[11] An increase in the rural population actually reduces government services per worker in the rural sector for two reasons: crowding of the population near the public good and an increase in the population that cannot easily utilize the good.

$$\tilde{g}_t = \frac{\xi\mu(1-\alpha)}{\alpha+\varepsilon\mu(1-\alpha)} \frac{F_t^{1-\varepsilon}}{M_t^{1-\xi}} g_t.$$

Note that the effective bias is influenced by the same parameters as the observed bias and, in addition, by the relative population size in the two sectors. We can examine the determinants of the effective bias further by writing out M_t and F_t in terms of demographic Eqs. (10.12) and (10.13), allowing us to rewrite the expression for the effective bias as

$$\tilde{g}_t = \frac{\xi\mu(1-\alpha)}{\alpha+\varepsilon\mu(1-\alpha)} \frac{[\varphi_t(1-\pi_{t-1})]^{1-\varepsilon}}{[\pi_{t-1}+(1-\varphi_t)(1-\pi_{t-1})]^{1-\xi}} \frac{g_t}{N_t^{\varepsilon-\xi}}.$$

For a given φ_t, a larger country population (N_t) increases the effective bias because of the advantage of sharing impure public goods in the city. Furthermore, one can show that an increase in N_t decreases φ_t as well, generating a further increase in the urban bias. Thus, as the population of the economy grows in size, the economy becomes more urbanized. During this "structural transformation," the *observed* bias in (10.15″) remains fixed but the *effective* bias, given above, widens.[12]

10.4 City Size and Development

In a recent article, J. Vernon Henderson (2010) discusses the connection between urbanization and growth. His article also contains interesting thoughts about strategies to deal with the rapid pace of urbanization in many of today's developing economies.

Consistent with Sect. 8.2, Henderson argues that while urbanization and growth have a strong positive correlation, there is not a simple causal effect running from urbanization to growth. The absence of a steadfast positive causal connection is most obvious in the situations where urbanization is actually associated with *falling* wages. This happens when migration to the city is driven by events that cause a drop in the relative productivity of rural workers. When workers are "pushed" to city there is a crowding of public infrastructure and government services that causes a decline in urban workers' productivity and in wages throughout the economy.

However, Henderson also argues that cities are associated with industry-level economies of scale and knowledge spillovers, features we have ignored, that can

[12]Our discussion suggests that an urban bias in the provision of effective government services is guaranteed. This is not the case because if both π_{t-1} and $1-\varphi_t$ are small enough then $[\varphi_t(1-\pi_{t-1})]^{1-\varepsilon}/[\pi_{t-1}+(1-\varphi_t)(1-\pi_{t-1})]^{1-\xi}$ could be large enough to make $\tilde{g}_t > g_t$. This situation is more relevant during the early stages of growth when urban crowding is not an issue and the city would have a small and slow growing population.

generate a rise in worker productivity. To capture these effects one could write the urban productivity index as

$$D_t = \left(\frac{G_t}{M_t}\right)^{\mu} E_t^{1-\mu} \bar{h}_t^{\sigma_1} M_t^{\sigma_2},$$

where σ_1 and σ_2 are positive parameters that capture productivity effects associated with living in a city. If we have human capital, as in some of the previous chapters, the term $\bar{h}_t^{\sigma_1}$ would capture the knowledge spillovers flowing from the average human capital of urban workers to the productivity of an individual worker. To capture economies of scale in production, or other urban agglomeration effects such as information flows or job matching, one could also include the term $M_t^{\sigma_2}$. An increase in the size of the urban workforce raises the productivity of the average urban worker.

There is empirical evidence indicating that greater industry size and a higher average human capital in a city causes individual worker productivity to be higher, other things constant. In addition to the advantage of sharing impure local public goods (see Sect. 8.3), these positive productivity effects within concentrated geographic areas explain why the modern sector firms tend to locate in cities. Agglomeration effects based on spatial concentration have the potential to create a causal connection running from urbanization to growth.[13]

Positive agglomeration effects suggest that there may be an optimal city size that maximizes worker productivity. Cities need a certain population size to fully exploit the industry economies and knowledge spillovers. However, these positive benefits may diminish at some point and the crowding of public infrastructure and services becomes an issue as the population continues to grow. Thus, weighing the benefits and costs of an increasing city population leads to the possibility of an optimal city size.

Empirical work by Henderson and his coauthors find that there is a concave, hump-shaped relationship between city population and individual worker productivity, confirming the existence of an optimal city size. They also find that the composition of industries within a city matters. The optimal city size is smaller the higher is the concentration of manufacturing industries. This is consistent with the fact that very large cities tend to have a high fraction of business and professional service industries.

The notion of an optimal city size is relevant in assessing the urbanization strategies in developing countries. Urbanization in today's developing world is both more rapid and more directed by national governments than it was in history. Historically, the typical process of urbanization, as a country's population goes from

[13]Duraton (2014) provides a discussion of the empirical evidence behind different mechanisms through which locating in cities can affect worker productivity. He also discusses the factors that limit worker productivity in developing countries, which is the focus of this chapter and the next section in particular.

roughly 15 to 75% urban, would take more than a century. In many developing countries, the same process of urbanization only takes a few decades.

The rapid pace of urbanization is one of the reasons that governments feel the need to intervene and control the process. China and India are considering using policy-based incentives to funnel their rural to urban migration into a relatively small number of mega-cities: 10–15 cities with average populations of 25–30 million people. This is quite controversial because the empirical work mentioned above suggests that the optimal size of city is much smaller. Thus, the formation of mega-cities may create a drag on aggregate growth.

10.5 Urbanization Today: New Mechanisms and Consequences

We would argue that most of the theory in this books applies equally well to historical and current development. However, as the last section reveals, there are at least some quantitative differences between historical and current development. Urbanization is occurring much more rapidly in today's developing countries than it did in history. It is also becoming clear that the current rapid rate of urbanization is caused by some forces that were not important in the past. In this section we discuss these new mechanisms for urbanization. The new mechanisms have important implications for the connection between agriculture and the traditional sector and the connection between urbanization and growth.

10.5.1 Consumption Cities: Urbanization Without Industrialization

In past history, and in many developing countries today, urbanization is closely associated with industrialization and an expanding manufacturing sector. Today, there is an important exception to this familiar pattern. Gollin et al. (2013) find that resource-exporters, e.g. oil exporting countries such as Gabon and Venezuela, have relatively high rates of urbanization despite having relatively low manufacturing sectors.

In the standard urbanization story, the population shifts from agriculture to manufacturing either due to the "pull" of increasing manufacturing wages or the "push" of declining agricultural wages. Instead, Gollin et al. focus on an income effect associated with a rise in rents from natural resources. They find that higher resource rents increase the demand for nontradable urban services—such as restaurant meals, cabs, apartments, and haircuts. The increase in demand raises the relative price of the nontradable services and causes a reallocation of labor to the city to work in the service sector. The reallocation of labor causes cities to grow near locations where the resources are extracted and refined. These are "consumption cities" because they are expanding due to an increase in demand for consumption services rather than an increase in the marginal productivity of urban labor generally.

Gollin et al. use a static four-sector model to articulate the mechanism. Manufacturing, agricultural, and resource goods (natural commodities) are traded in international markets that set their prices. Manufacturing goods and urban services are produced in cities, while agricultural and resource goods are produced in the rural sector. Urban services are nontraded, so their price is determined by domestic market conditions. A rise in the international price of resource goods raises income for countries that export these goods. The increase in income raises demand for all goods, but only the price of urban services rise because other prices are determined internationally. The rise in the relative price of urban services raises the relative wage in the service sector and pulls labor out of other sectors. Even though employment in the manufacturing sector declines, total urban employment can rise because labor is also being pulled out of agriculture. In this situation, an increase in urbanization is associated with a decline in the manufacturing sector.

10.5.2 Mushroom Cities: The Traditional Sector Grows in the City

Another difference between the past and the present is the degree to which urbanization results from population growth *within* cities—what is called "urban natural increase." Jedwab et al. (2014) show that the rapid urbanization of today's developing countries is driven more by natural increase than by the rural-to-urban migration of European history. The urban centers of Europe during industrialization were "killer cities"—with high mortality and low fertility. In contrast, developing countries today largely contain "mushroom cities"—with low mortality and high fertility. They show the resulting difference in the urban natural increase across these two types of cities accounts for most of the difference in the current and historic urbanization rates.

Jedwab et al. also establish that the rapid growth in cities today is correlated with a high proportion of urban residents living in slums, low investment in urban human capital, and a larger informal sector. This scenario suggests that cities in today's developing countries contain a sizeable traditional sector, with households that produce informally, have many children, and do not investment much in education. In the face of the rapid urban natural increase, rural households still continue to move to the city. However, the gaps in living standards between rural households and those living in slums are likely to be small. These household share more similarities than differences. It is best to think of them both as residing and producing in the traditional sector.

10.5.3 Urbanization Without Growth Revisited

Today's developing countries are urbanizing at historically rapid rates. However, their growth rates in worker productivity are, as a group, no different than those of currently rich countries. There is no convergence for most poor countries today (Chap. 7).

The combination of unusually rapid urbanization and mediocre economic growth rates raises doubts about viewing urbanization as an engine of growth. The lack of correlation between growth and urbanization could be due to the weak connection between urbanization and manufacturing for today's developing countries. Today's urbanization is associated more with growing service and informal sectors rather than a robustly expanding manufacturing sector. This pattern also implies a break in the close connection between the urban sector and the modern sector, as today's developing cities often have large slums that have more characteristics in common with the traditional sector.

10.5.4 Slums

A consistently disturbing feature of early urbanization, whether in a historical or contemporary development setting, are areas of poor quality housing and inadequate basic public services, commonly known as "slums." The problem is more severe today than in the past. Currently, 860 million people are living in slums around the world (Marx et al. (2013)). Slums are the most dramatic indicators that the crowding of public services in urban areas is a major problem of development. As the theory suggests, if urbanization is predominantly driven by population growth and dire conditions in rural areas, there will be a decline in urban services per worker and wages will fall throughout the economy.

Despite the deplorable conditions in slums, urban poor world-wide are richer and happier than those that remain in rural areas, at least on average (Glaeser 2011). This suggests just how desperate life must be in the country-side of many developing countries. In this situation, aid directed at developing public infrastructure in rural communities would increase productivity throughout the economy by reducing migration flows to the city and raising urban public services per worker.

Marx et al. (2013) argue that slums can cause problems that policy makers do not fully account for. They are particularly concerned with the possibility that slums create intergenerational poverty traps. One can easily see how this could happen. Moving to the city raises wages and family resources. However, living in the slum is a less healthy environment because of the reduced space and overwhelmed public infrastructure that results in generally unsanitary conditions. While, as reported in Chap. 5, the health conditions in rural areas are generally not as good as in cities, there may be city slums where this is not true.

Young children in the family are most susceptible to the disease and illness that an unhealthy environment breeds. Their compromised health could have long-term consequences for their adult productivity because of stunted physical development and impaired ability to learn. While the family as a whole has greater resources in the short-run, the next generation may be less productive than workers raised in rural areas. In short, the move to the city was bad for the future generations of family. This could be true even if parents possess intergenerational altruism because the short-term benefits to the current generation of the family may be greater than the cost to the next generation.

An intergenerational poverty trap of this type could easily be ignored by urban elite policy makers. Some aspects of urban crowding, such as crowding of roads and energy provision, lower the productivity of all urban workers. However, if the urban elite live a safe distance from the slums, they will not be exposed to the unhealthy environment created by a crowding of health and sanitation services. As a result, the policy makers will underestimate the full cost of migration to the city and the full benefit of basic public services provided to the slums. Even if the goal is to maximize aggregate economic growth, there will be too little public investment in rural areas and the composition of urban public spending will allocate too few resources to the slums.

10.6 Hukou

Chinese policy provides an interesting example of internal migration restrictions (Meng 2012; Wang and Piesse 2010; Wang and Weaver 2013). When the Chinese Communist Party rose to power in 1949, no labor mobility between rural and urban sectors was allowed. Rural to urban migration was formally prevented starting in 1955 by a household registration system, known as "Hukou," that established household residency. Those with urban Hukou were given access to high paying jobs and government services. The advantage of rural Hukou was a claim to land use. During the first 30 years of Communist rule, more than 80% of the population remained in rural areas.

As the Chinese economy was reformed and began to grow in the last quarter of the twentieth century, migration restrictions were relaxed due to the need for labor in the urban manufacturing sector. However, it has been uncommon for a rural person to be granted a permanent urban Hukou. Most migrants to the cities are younger workers that live and work in the cities on a "temporary basis," without access to the government services enjoyed by urban residents. Between 1990 and 1997, rural migrants working in cities increased from 25 to 37 million. By 2009, the number of rural migrant workers quadrupled to reach 145 million.

Due to the strong demand for labor, unemployment for the migrant workers is low and hours worked are high. Migrants work over 60 h per week, which is likely much more than they would have worked in rural settings. The expansion in labor supply and the reallocation of workers to a sector where they are more productive per hour are important reasons for China's continued strong economic growth.

Despite the migration of these temporary workers, large wage gaps remain. There was a three-fold per capita income gap favoring urban residents at the end of the twentieth century. The significant work opportunities for the migrants and the large wage gaps both indicate that Chinese labor markets suffer from significant inefficiency due to the persistence of mobility restrictions.

Recently there has been a movement in China to reduce internal migration restrictions and increase the population flow to cities. The primary motivation for the urbanization push is to increase domestic demand for marketed goods. The urbanization push coincides with a desire to eliminate the Hukou system. This raises difficult questions about how deal with the land rights of potential rural migrants. In some cases rural residents are forced off their land or given far below market value for the transfer of land to local government authorities. There have been significant charges of "land grabs" by local officials that benefit from reselling the land at market value.

Access to the work opportunities in the city may be enough to compensate younger workers for the loss in their claims to land. It is less clear that this is true for older workers, especially because there is uncertainty about access to government services and pension for the new urban residents. The large migration to the city is straining the capacities of governments to provide the services and benefits that previous urban residents have enjoyed.

10.7 De-urbanization: Past and Present

Many view England, generally recognized as the place where the Industrialization Revolution began, as the nation that was the home of the first modern economy. However, as carefully documented by de Vries and van der Woude (1997), the first modern economy with sustained economic growth may have instead been developed in the Netherlands during the sixteenth and seventeenth centuries. The Dutch economy was not modern in the sense of being industrial but it was modern in the sense of being built on commerce and trading in domestic and international markets. The early appearance of commerce was fostered by high levels of education in the population, a monetary system, and well defined and enforced property rights.

Another advanced feature of the Netherland economy was the high percentage of the population living in urban areas. Urbanization was rapid over the sixteenth and seventeenth centuries, followed then by a period of de-urbanization:

1525	31.5%
1675	45%
1750	42%
1795	40%
1815	38%

An urbanization rate of 45% made the Netherlands far and away the most urbanized country in Europe at the end of the seventeenth century. In comparison, England did not achieve an urbanization rate of 40% until the middle of the nineteenth century. The high rate of urbanization in the Netherlands was associated with a structural transformation of their early economy from one based on traditional agriculture to one based on shipping, ship building, whaling and fishing, brewing, and textiles.

However, what explains the period of de-urbanization that followed? de Vries and van der Woude suggest three reasons. First, exports fell sharply after 1700. This was due to trade wars, in some cases leading to actual wars, and due to the fact that the Dutch began to lose their comparative advantage in ship building and textiles. The Dutch lost their leadership position in these areas as new technologies were being developed in other European countries. Second, the early Industrialization

Revolution created a proto-industry in rural areas. Textiles were produced by women and children out of ordinary homes rather than by the more expensive urban craftsmen. Finally, the cities suffered from crowding that created hygienic problems and disease.

The Dutch experience with early urbanization provides indirect evidence for the view that urbanization, especially early in development, is more a byproduct of growth rather than a cause of it. When the underlying determinants of growth falter, the wage differential between urban and rural areas shrinks and no longer provides sufficient compensation for the problems associated with urban crowding. As a consequence urbanization slows or is reversed.

A similar de-urbanization is being forecasted for India today. Just as in many developing countries, India has experienced rapid urbanization since WWII, although about half of its workforce is still in agriculture. There are signs that the labor flow to the cities has now reversed. Growth in the urban manufacturing sector is slowing due to crowding and insufficient infrastructure and due to the high cost of complying with a myriad of labor laws. Crisil, a large economic and business research group headquartered in Mumbai, predicts that a reverse migration of 12 million workers out of the cities and back to relatively low-wage agriculture will take place between 2013 and 2019. Just as in the Dutch Republic of the eighteenth century, if the conditions sustaining economic growth are compromised, and urban crowding is not dealt with, de-urbanization can occur. Urbanization per se is not a sufficiently strong engine for growth to sustain itself.

10.8 Conclusion

Many observers of fiscal policy in developing countries see large disparities in the government's provision of goods and services across regions in favor of urban areas. The most common interpretation of these disparities is that urban households have disproportionate political power that causes fiscal policy to favor them at the expense of rural households. This chapter develops a simple theory of urban bias in the allocation of productive government services suggesting that the common interpretation may not be correct. An urban bias in the provision of government services can be explained by efficiency considerations and it may be the case that the extent of the bias is entirely independent of the weight that policymakers place on urban and rural households. The observed inequities in the provision of government services across rural and urban areas are not necessarily the result of a redistributive urban bias in the politics of developing countries, but rather serve to raise wages throughout the economy.

Our analysis indicates that urban-rural politics is more likely to influence policies that directly affect the cost of migration from rural to urban areas. An increase in the costs associated with rural to urban migration does clearly raise the welfare of urban households at the expense of rural households.

We ignored the intergenerational consequences associated with living in the unhealthy slums of cities in developing countries. Intergenerational consequences of unhealthy slums could create poverty traps that slow economic growth and are worthy of future research. To address the possibility at least three new features will have to be introduced into the model. First, there must be two types of urban households: one type lives in the slums and one type does not, the urban elite who set the policy. Second, there must be two types of public services: those that affect the productivity of all urban workers (roads) and those that affect only those living in the slums (public health and sanitation services). Third, the public services provided in the slums must influence the adult human capital of the children living there.

10.9 Exercises

Questions

1. How are the traditional and modern sectors interpreted in this chapter? What is difference between sectors?
2. What is the source of the wage gap between sectors? Write down and explain the equation for the wage gap.
3. Describe the government's behavior. How are the government's preferences modeled? What are the government's constraints and choice variables?
4. Explain why there is an urban bias in government service provision. Why is the urban bias efficient?
5. What is the role of free migration of households across sectors in establishing the efficient urban bias? What happens if there is no migration across sectors?
6. What are "push" and "pull" factors that create migration to the cities? Explain how these factors affect government service provision.
7. How does balanced technological progress across sectors affect the following variables? What happens if the technological progress only occurs in the urban sector?
   (a) urbanization
   (b) government service provision
   (c) wages in each sector
   (d) wage inequality
8. What is a redistributive urban bias? How does an increase in migration restrictions affect the variables in *Question* 7?
9. If the government provides transfers in addition to productive government services is there still an efficient urban bias?
10. . How does accounting for endogenous tax revenue affect the urban bias?
11. If productive government services are provided by impure public goods, such as roads, how does it affect the urban bias?
12. Describe how urbanization in today's developing world differs from the historical urbanization of the U.S. and Europe. What implications does this have application of the dual economy model? For government policy?

13. How can intergenerational effects of living in unhealthy slums cause the allocation of productive government services to rural areas to be lower than the allocation that would maximize economic growth?
14. Discuss some of the issues related to urbanization in China.
15. Make a case for the idea that urbanization causes economic growth. Next, make a case for the idea that urbanization has a neutral or even negative causal effect on economic growth.
16. Relate the concept of "consumption cities" to the McMillan-Rodrik explanation for the difference between worker productivity growth rates in Asia and Africa (Chap. 7).

Problems

1. Show that $\pi_t = \pi_{t-1} + (1 - \varphi_t)(1 - \pi_{t-1})$ and $1 - \pi_t = \varphi_t(1 - \pi_{t-1})$.
2. Show that one can use (10.6b), along with (10.2), (10.5), and (10.13), to derive the following expression for the fraction of the traditional population that remains in the traditional sector,

$$\varphi_t = \left\{\frac{\left[\left(\tilde{E}_t/E_t\right)^{1-\mu}\left(\tilde{g}_t\right)^{\mu}\right]^{1-\alpha}}{k^{\alpha} g_t^{\mu}\left(\tilde{W}_t/W_t\right)}\right\}^{1/\alpha}\frac{L}{n\tilde{N}_{t-1}E_t}.$$ Use this equation to show $$\frac{\partial \varphi_t}{\partial \tilde{g}_t} = \frac{\mu(1-\alpha)}{\alpha}\frac{\varphi_t}{\tilde{g}_t}.$$

3. Using the results of *Problems* 1 and 2, choose $\tilde{g}_t$ and g_t to maximize $V_t((1-\tau)W_t) + \upsilon V_t\left((1-\tau)\tilde{W}_t\right)$ subject to $g_t\pi_t + \tilde{g}_t(1-\pi_t) = b_t$. The first-order condition associated with the choice of $\tilde{g}_t$ can be used to derive the efficient urban bias equation given in (10.15).
4. Use (10.13) to argue that technological progress lowers wage inequality.
5. Use (10.2), (10.3b), (10.13), (10.15), and the government budget constraint, $g_t\pi_t + \tilde{g}_t(1-\pi_t) = b_t$, to argue that migration to the city resulting from push factors, for a given value of E, will reduce wages across the economy.
6. *Extensions*. Write down and carefully explain the urban bias condition when
   (a) the government provides consumption goods and services or transfers in addition to productive services.
   (b) tax revenue endogenously depends on the share of the population in the urban area.
   (c) productive government inputs are impure public goods.
7. In the case where government productive inputs are impure public goods, explain the difference between the "observed" and the "effective" urban bias. How is each affected by population growth?

References

Bates R (1981) Markets and states in tropical Africa. University of California Press, Berkeley.

Bezemer D, Headey D (2008) Agriculture, development, and urban bias. World development 36 (8):1342–1364.

Black D, Henderson V (1999) The theory of urban growth. Journal of political economy 107 (2) 252–284.

Bloom D, Khanna T (2007) Urban revolution. Finance and development 44(3):8–14.

de Vries J, van der Woude A (1997) The first modern economy. Cambridge University Press, New York.

Duranton G (2008) From cities to productivity and growth in developing countries. Canadian journal of economics 41(3):689–736.

Duraton G (2014) Growing through cities in developing countries. Policy research working Paper 6818. World Bank, Washington, DC.

Duranton G, Puga D (2004) Micro-foundations of urban agglomeration economies. In: Henderson V, Thisse J (eds) Handbook of regional science and urban economics, vol. 4. North-Holland, Amsterdam.

Fay M, Opal C (2000) Urbanization without growth: A not so uncommon phenomenon. World Bank working paper 2412. World bank, Washington, DC.

Feler M, Henderson V (2011) Exclusionary policies in urban development: Underservicing migrant households in Brazilian cities. Journal of urban economics 69(3):253–272.

Glaeser E (2011) Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier. Penguin Press, New York.

Gollin D, Jedwab R, Vollrath D (2013) Urbanization with and without industrialization. University of Houston Department of Economics working paper 2013-290-26.

Hansen G, Prescott E (2002) Malthus to Solow. American economic review 92(4):1205–1217.

Harris R, Todaro R (1970) Migration, unemployment, and development: A two-sector analysis. American economic review 60(1):126–142.

Henderson V (1974) The sizes and types of cities. American economic review 64(4):640–650.

Henderson V (2010) Cities and development. Journal of regional science 50(1):515–540.

Henderson V, Wang HG (2005) Aspects of rural-urban transformation of countries. Journal of economic geography 5(1):23–42.

Jedwab R, Christiaensen L, Gindelsky M (2014) Rural push, urban pull and. . .urban push? New historical evidence from developing countries. George Washington University Institute for International Economic Policy working paper 2014–04.

Lipton M (1977) Why poor people stay poor: Urban bias in world development. Harvard University Press, Cambridge, Massachusetts.

Liu Z (2014) Human capital externalitiesmin cities: evidence from Chinese manufacturing firms. Journal of economic geography 14:621–649.

Lucas R (2004) Life earnings and rural-urban migration. Journal of political economy 112(1):S29–S59.

Majumdar S, Mani A, Mukand S (2004) Politics, information, and the urban bias. Journal of development economics 75(1):137–165.

Marx B, Stoker T, Suri T (2013) The economics of slums in the developing world. Journal of economic perspective 27(4):187–210.

McCormick B, Wahba J (2003) Did public wage premiums fuel agglomeration in LDCs? Journal of development economics 70(2):349–37.

Meng X (2012) Labor market outcomes and reforms in China. Journal of economic perspectives 26 (4):75–102.

Moretti E (2004a) Human capital externalities in cities. In: Henderson V, Thisse JF (eds) Handbook of regional and urban economics, vol. 4. North Holland, Amsterdam, pp 2243–2291.

Moretti E (2004b) Worker's education, spillovers, and productivity: Evidence from plant-level production functions. American economic review 94(3):656–690.

Mourmouras A, Rangazas P (2013) Efficient urban bias. Journal of economic geography 13 (3):451–471.
Rangazas, P, Wang, X (2018) Internal migration restrictions and labor allocation in developing countries. Singapore economic review
Rosenthal S, Strange W (2004) Evidence on the nature and sources of agglomeration economies. In: Henderson V, Thisse J (eds) Handbook of regional and urban economics, vol. 4. North Holland, Amsterdam.
Wang X, Piesse J (2010) Inequality and the urban-rural divide in China: Effects of regressive taxation. China and the world economy 18(6):36–55.
Wang X, Weaver N (2013) Surplus labor and urbanization in China. Eurasian economic review 3 (1):84–97.

11 Conclusion

In this chapter we summarize some of the main points we have learned about development and their related policy implications. The points are organized into those related to the *onset of growth* and those related to the *nature of growth* once it begins. We also discuss policy implications and topics for future research.

11.1 The Onset of Growth

We have seen that sustained growth is a very recent phenomena, starting two or three centuries ago for some countries and within the last few decades for many others. The very different timing of the onset of modern growth is the main reason for the divergent standards of living across the world today. This is particularly true because average growth rates remain trendless for long-periods of time as economies make the slow transition toward their steady states. With growth rates that are similar at all stages of development, it is not generally true that poorer countries converge or catch up to richer countries. The features of economies that determine the timing of the initial growth take-off are then crucial.

11.1.1 The Appearance of a Modern Sector

In Chap. 6 we present a theory that helps us understand the conditions giving rise to the production in factories that sets the stage for modern growth via the accumulation of physical capital. Before these conditions are met, we can think of the economy as being trapped in a situation where goods are produced using informal traditional methods. The key consideration is whether factories can compete profitably with traditional methods while paying workers at least as much as they make in the traditional sector and capital owners at least what they could earn by owning land. For a given state of technology, the following conditions increase the likelihood that profitable factories will appear.

S. Das et al., *Economic Growth and Development*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-319-89755-4_11

1. *Limited Natural Resources*—Traditional methods rely heavily on direct use of natural resources. Countries that have an abundance of usable land and other easily accessible natural resources will have a productive traditional sector. This leads to relatively high wages and/or high returns to the ownership of the natural resources. High factor prices paid to labor and land means operating factories will be costly. Countries without these resource advantages have cheaper labor and lower required returns to assets that paves the way for a profitable modern sector.
2. *Limited Political Power of Landowners and Traditional Craftsman*—Landowners and craftsmen oppose the competition from factories and seek to prevent their appearance. They support policies that favor the traditional sector and block the formation of large firms by legislating various restrictions, taxes, and regulations in the modern sector. Pro-growth dictators or democratic participation of the general population limit the influence of large landowners and established craftsmen and thereby encourage factory based production and physical capital formation.
3. *Legal Institutions Protecting Property Rights*—Traditional production relies on informal arrangements between workers, landowners, craftsmen, and customers. In local settings where trading occurs on a small scale between parties that know each other well, these informal arrangements suffice. However, to produce and trade on a larger and more impersonal manner, a legal system that protects property rights is critical substitute for the trust and local information that is present in traditional settings.
4. *Public Infrastructure*—Mechanized factory production and trading at a larger scale requires the provision of public capital inputs such as utilities and roads.

Equation (6.8) from Chap. 6 provides the profit requirement for the appearance of the modern sector. The conditions listed above can be related to Eq. (6.8) as follows. Condition (1) determines the value of $\tilde{A}l^{\alpha}$ in the traditional sector, while (2)–(4) determine the value of A in the modern sector. Given the state of available technology and the country's natural resources, the key is to generate a high level of A. There is no unique way to accomplish this. For example, excellent roads and strong property right protection can offset anti-capital policies supported by large landowners. The sources of A to focus on, in order to get modern growth going, are likely to be country specific; depending on geography, culture, and political realities.

11.1.2 Poverty Traps and Schooling

Even after modern growth begins, the theory used in this text suggests that schooling of older children may not take off. The schooling poverty trap from Chaps. 5 and 6 is independent of the technological change and capital accumulation associated with modern growth. There is evidence supporting this possibility, as many countries initially experience sustained positive growth without seeing much progress in schooling and literacy across the workforce. The schooling poverty trap is most persistence when (i) there are no significant noneconomic reasons to educate the general population, (ii) the relative productivity of children is high, and (iii) there is a lack of government effort to encourage mass education.

11.2 The Nature of Modern Growth

The theory in Chaps. 7, 8, and 9 identifies and explains some common features of growth as economies begin to develop and modernize. The early period of modern growth is dominated by the structural transformation of the economy as labor moves out of the traditional sector and into the modern sector. The sources of the growth in worker productivity during this initial phase are the underlying technological progress, the expansion in work hours, and the aging of the work force as fertility falls. Work hours expand because the opportunity for working long hours throughout the year is greater in the factories of the modern sector and because families that relocate to the modern sector have fewer children, which allows more time for market work. The decline in fertility also reduces the number of young workers and causes the average age and productivity of the workforce to increase. In some settings, where labor markets are particularly inefficient, there may also be gaps in productivity per hour worked that favor the modern sector. These gaps create an additional source of growth as labor migrates to the modern sector during the structural transformation.

During the early stages of the structural transformation, the accumulation of human and physical capital often does not play a large role in economic growth. The structural transformation itself creates offsetting effects on physical capital intensity in the modern sector. Saving rates rise and fertility falls, which raises capital intensity, but the movement of labor toward the modern sector causes a crowding of the private capital and reduces capital intensity. Schooling often is trapped at low levels or increases very slowly. One reason that some countries created very high growth rates after World War II is that their policies pushed the development of human and physical capital during the early stages of the structural transformation.

In the later stages of development, as fertility continues to fall and the pace of the migration to the modern sector slows, physical capital intensity rises. In addition, schooling of older children accelerates and creates further downward pressure on fertility. During this stage, the growth rates in output per worker increase and then stabilize for many decades. The rise in physical and human capital per worker also cause significant increases in output per hour worked. The long transition, at relatively trend-less growth rates, implies that convergence of income across countries can be slow unless lagging countries adopt unusually aggressive pro-growth policies.

The growth over this long transitional period is fueled by events that must eventually come to an end. The end of the structural transformation, the leveling off of the fertility decline, and the prosperity associated with the rise in productivity per hour worked, all cause work hours to fall rather than rise. The rise in the rates of investment in physical and human capital also levels off, leaving the diminishing returns to investment to dominate. These events create growth slowdowns that can only be offset by an increase in technological progress. The prospect of a general growth slowdown, especially for the most developed economies, is a major concern for the twenty-first century.

11.3 Policy Implications

In this section we summarize some of the main policy implications suggested by the theory.

11.3.1 Domestic Growth Policy

In discussing the conditions needed for the appearance of factory-based production, we mentioned the importance of limiting the influence of those with vested interest in the traditional sector, a legal system that supports formal trading in large markets, and public infrastructure. Of these key elements, we have primarily focused on the importance of investment in productive public infrastructure. For simplicity, many researchers choose to combine public capital, such as roads, with private capital into a single measure of the economy's physical capital stock. This is a very misleading modeling strategy. As in Chaps. 5, 8, and 10, public capital should be modeled as a complement rather than a substitute for private capital. Public capital is an essential input to production that raises the marginal productivity of private capital. This is why growth in the public capital stock is needed to raise TFP, as traditionally measured, paving the way for the emergence of profitable private capital and the modern sector.

As emphasized in Chap. 8, an important byproduct of the structural transformation is that economic activity in the modern sector is easier to tax. Thus, the structural transformation naturally increases the economy's capacity to raise tax revenue. An important determinant of the pace of growth is how governments use this capacity. Countries that use the increased revenue for public sector investment will grow faster than those that use it for government consumption and transfers that finance private consumption. Growing governments are often seen as a drag on the economy. However, growth rates need not fall as governments become larger if a significant portion of the increased tax revenue is invested.

Chapter 10 suggests that the *allocation* of public investment is almost as important as the *level* of public investment. As with private capital formation, internal migration to the modern sector crowds the supporting public capital infrastructure. It is optimal for governments to control the pace of migration by investing in public capital that raises productivity and living standards in the traditional sector as well as well as the modern sector. Chapter 10 argues that there is an efficient mix of public investments across sectors that maximizes wages throughout the economy. For various reasons, the efficient mix is biased toward the modern sector, but the traditional sector cannot be ignored.

Apart from public infrastructure investment, in Chaps. 5, 7, and 9 we stress that education of the workforce is critical for accelerating and sustaining economic growth. There is the direct effect of education on the productivity of workers throughout the economy, but there are important indirect effects as well. Increases in human capital speed the structural transformation and reduce fertility, both of which accelerate growth.

We have highlighted the possibility of poverty traps associated with schooling. In environments that generate schooling traps, government action is needed for human capital growth to take place. There are various ways to jump-start the growth, from child labor and compulsory schooling laws to subsidies that replace a portion of the forgone earnings lost to the family when older children attend school. The key is to get schooling going because rising schooling increases the likelihood that schooling will continue to rise, even in the absence of government policies and laws.

11.3.2 International Trade, Capital Mobility, and Foreign Aid

Throughout the text we have discussed how various types of international activity affect economic growth. In Chap. 5, we found that opening the economy to physical capital inflows significantly speeds up transitional growth. It also raises the effective saving rate of low saving countries leading to greater steady state capital formation and worker productivity. In addition, opening the economy to capital flows affects the structure of fiscal policy, creating incentives for governments to lower tax rates and to invest a greater fraction of tax revenue in productive public capital. The pro-growth change in fiscal policy provides another reason that opening the economy raises worker productivity and living standards in the long-run.

In Chaps. 6 and 8, we found that opening the economy to the international trade of goods can slow the structural transformation for developing economies that have a comparative advantage in the goods produced in the traditional sector. As we discussed at length, there are various reasons why slowing the structural transformation may lower aggregate economic growth. However, while it is possible that aggregate growth may be slowed, the majority of the population may nevertheless be made better-off by the international trade of goods. This is most likely when the growth effects of opening are small, due to the presence of important *domestic* impediments to growth, and when the benefits of growth tend to be concentrated in a relatively small segment of the population.

In Chap. 5 we examined when foreign aid would most likely raise growth and long-run worker productivity. Unconditional budget aid to countries with poor growth records has weak short-run growth effects and no long-run effects on worker productivity. On the other hand, forcing pro-growth fiscal reform is too costly in terms of compensating aid inflows and is likely to be circumvented by uncooperative governments. Aid is more likely to work when the conditions of aid are negotiated and owned by pro-growth governments that have specific ideas about investment projects that are likely to work best in their country.

11.4 Ideas for Future Research

We view the book as an introduction to the integration of the new features of growth theory into a two sector framework. We hope it provides a foundation for doing further reading and conducting future research on topics where a two-sector approach is important.

Fundamental sources of differences in schooling across traditional and modern sectors should be more thoroughly examined, including the possibility of cultural differences that affect expectations about the benefits of human capital investments. There is evidence of community –level human capital externalities which could reflect the importance of cultural attitudes and informed expectations about the benefits of education. A more detailed analysis of what policies are most effective in raising schooling levels, especially in traditional settings, is needed. We also need a better understanding of how the structural transformation affects average schooling levels in the economy. Does the move to the modern sector generally raise human capital investment in health and education?

Additional work is also needed to identify whether the tradition of keeping land within the family is due to intergenerational preferences or due the absence of land markets. If the absence of land markets is an important factor that binds labor to the traditional sector, what does it take for land markets to arise and how can the government promote their development?

A more complete study of the structural transformation should include a special focus on the large slums that form in the cities of developing countries. The possibility that slums create poverty traps, associated with poor health environments that impair the development of children, is an important area for further research.

Given concerns about urban crowding, the optimal pace of the structural transformation should be pinned down more precisely. What should the government do to speed up or slow down the structural transformation? Are industrial policies that speed up the structural transformation justified? Should they subsidize all modern industries or are there valid reasons to focus on particular industries? When, if ever, does it make sense to use migration restrictions or agricultural policy to slow down the structural transformation?

We hope that the book has made a case for examining the effects of international trade in dynamic two-sector models. The effects of trade on growth and the welfare of different household types and generations are complex. It is fair to say that we are far from any definitive understanding in this area. The strong case for free trade is predominately driven by results from static models and we need to examine more carefully when trade is beneficial to developing countries in models that incorporate growth and intergenerational welfare effects.

Technical Appendix

This appendix gives a quick refresher of the topics in college algebra and basic calculus, and their extension to optimization theory, that are used in the models of the text. To see the different concepts in action, we have included EXAMPLES FROM THE TEXT as each topic is reviewed. More advanced mathematical methods are found in the chapter appendices and references.

A.1. Two Useful Functions

We use two types of functions frequently in the text.

Power Functions

A power function has the general form

$$y = f(x) = Ax^a,$$

where x is a nonnegative variable and a and $A$are positive constants, or *parameters*. In words, the function says that y is an increasing function of x, but the relationship between the two variables can have a variety of characteristics depending on the precise value of a.

For,

$0 < a < 1$	$f(x)$ is a concave function of x
$a = 1$	$f(x)$ is a linear function of x
$a > 1$	$f(x)$ is a convex function of x

If these shapes are not familiar, set $A = 1$ and plot $f(x)$ for different values of x, given a value of a that satisfies each of the three different cases above.

When dealing with power functions you need to remember some of the algebra associated with expressions that are raised to a power. Here are some important algebraic results for such expressions, where a and b are positive parameters.

S. Das et al., *Economic Growth and Development*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-319-89755-4

(i) $x^a x^b = x^{a+b}$
(ii) $(x^a)^b = x^{ab}$
(iii) $(x^a)^{1/b} = x^{a/b}$
(iv) $x^{-a} = \frac{1}{x^a}$
(v) $\frac{x^a}{x^b} = x^a x^{-b} = x^{a-b}$
(vi) $(xz)^a = x^a z^a$, where z is another nonnegative variable

We also need to remember the rules for differentiating a power function with respect to x.

(i) First derivative

$$f'(x) = aAx^{a-1}$$

(ii) Second Derivative

$$f''(x) = (a-1)aAx^{a-2}$$

Note that, given our assumptions, $f'(x) > 0$. However, the sign of $f''(x)$ depends on the precise value of a. The sign of the second derivative is important because it offers a way of identifying the shape of the function without the need to form plots.

For,
$0 < a < 1$ $f''(x) < 0 \Rightarrow f(x)$ is a concave function of x
$a = 1$ $f''(x) = 0 \Rightarrow f(x)$ is a linear function of x
$a > 1$ $f''(x) > 0 \Rightarrow f(x)$ is a convex function of x

A way of understanding the connection between the second derivative and the shape of $f(x)$ is to note that the second derivative tells us what is happening to the *slope* of $f(x)$, i.e. it gives us the change in the first derivative when there is an increase in x.

For,
$0 < a < 1$ $f'(x)$ is *falling* as x increases, so the graph becomes flatter
$a = 1$ $f'(x)$ is *constant* as x increases, so the graph remains linear
$a > 1$ $f'(x)$ is *increasing* as x increases, so the graph becomes steeper

EXAMPLES FROM THE TEXT

In Sect. 2.1 of Chap. 2, we find the following power function representing total production,

$$Y_t = AK_t^{\alpha} M_t^{1-\alpha}.$$

Using the algebra associated with variables raised to exponents, production can be written on a per worker basis, *worker productivity*,

$$y_t \equiv \frac{Y_t}{M_t} = \frac{AK_t^{\alpha} M_t^{1-\alpha}}{M_t} = \frac{AK_t^{\alpha} M_t M_t^{-\alpha}}{M_t} = \frac{AK_t^{\alpha}}{M_t^{\alpha}} = Ak_t^{\alpha},$$

where $k_t \equiv \frac{K_t}{M_t}$.

The first and second derivatives of the worker productivity function with respect to k_t are $\alpha A k_t^{\alpha-1} > 0$ and $\alpha(\alpha - 1)Ak_t^{\alpha-2} < 0$. Worker productivity is an increasing concave function of k_t.

(Natural) Logarithmic Function

Our other special function is the natural logarithmic function, which we refer to as just the log function. The log function is an increasing concave function of the form,

$$y = f(x) = A \ln x,$$

where x is a positive variable and A is a positive parameter. As with the power function, if you are not familiar with the shape of the log function you should set $A = 1$ and plot the function for different values of x.

Alternatively, we can learn about its shape by recalling the rules of differentiation for log functions,

(i) First Derivative

$$f'(\mathrm{x}) = \frac{A}{x} > 0$$

(ii) Second Derivative

$$f''(\mathrm{x}) = -\frac{A}{x^2} < 0$$

As with the power function when $a < 1$, the derivative of the logarithmic function is positive and decreasing as x increases, i.e. its slope becomes flatter at higher values of x.

The following results will be useful when doing algebra with expressions involving logs. The parameter a and the variable z are both positive values.

(i) $\ln(xz) = \ln x + \ln z$
(ii) $\ln\left(\frac{x}{z}\right) = \ln x - \ln z$
(iii) $\ln(x^a) = a \ln x$

EXAMPLES FROM THE TEXT

The single period utility function we use throughout the text takes the natural log form, $u = \ln c$. The marginal utility of consumption is the derivative of u with respect to c, $1/c > 0$. The marginal utility of consumption is clearly decreasing in c. This can also be verified by taking the second derivative with respect to c, $-1/c^2$, which tells us how the marginal utility of consumption changes with c.

In Sect. 4.1 of Chap. 4, deriving the optimal choices for consumption, fertility (n_{t+1}), and schooling (e_t) is simplified by using the algebraic rules for taking the natural log of a product. The extended utility function in Chap. 4 is

$$U_t = \ln c_{1t} + \beta \ln c_{2t+1} + \psi \ln\left(n_{t+1} h_{t+1} w_{t+1} D_{t+1}\right),$$

where $h_{t+1} = e_t^\theta$ is the human capital production function relating parent's choice of schooling time for children to the resulting human capital when the child becomes an adult. The expression for the parent's lifetime utility can be written as

$$\begin{aligned}&\ln c_{1t} + \beta \ln c_{2t+1} + \psi \ln n_{t+1} + \psi \ln e_t^\theta + \psi \ln\left(w_{t+1} D_{t+1}\right) = \\ &\ln c_{1t} + \beta \ln c_{2t+1} + \psi \ln n_{t+1} + \psi\theta \ln e_t + \psi \ln\left(w_{t+1} D_{t+1}\right).\end{aligned}$$

Note, when maximizing U_t to find the optimal household behavior, that the last term above is unaffected by household choice. The derivatives representing the marginal utility of household choices are $1/c_{1t}$, β/c_{2t+1}, ψ/n_{t+1} and $\psi\theta/e_t$.

A.2. Optimization

Single Choice Variable

The two special functions discussed in the previous section are increasing in x. This means that they have no maximum value. In economic terms, if these functions represent output or utility, as x increases there is always a marginal benefit. However, because of scarcity, there is typically also a cost to increasing x. For simplicity, suppose the scarcity is reflected in the fact that sellers of x charge a price, p, for its use. Also assume the market for x is competitive, so individual agents take the value of p as given (unaffected by their actions)

The *rationality* assumption in neoclassical economics says that agents will assess both the benefits and costs of making a decision and make choices that do not systematically deviate from the choice that maximizes the net benefit. To illustrate how this assumption works, we create a new function that reflects both the benefit and the cost of choosing x. The simplest function that illustrates this idea is the *profit function*. Let the profit function be defined as,

$$\tilde{f}(x) = Ax^a - px,$$

where $A > 0$, $p > 0$, and $0 < a < 1$. The first and second derivatives of the profit function are

$$\tilde{f}'(x) = aAx^{a-1} - p$$

$$\tilde{f}''(x) = (a-1)aAx^{a-2} < 0.$$

Note that the second derivative is negative, so the profit function is concave. This also tells us that the first derivative is decreasing. However, the first derivative can have any sign. When x is low it is more likely to be positive. A positive derivative indicates that total profit increases as x increases. As x increases the value of the first derivative falls, the marginal profit becomes smaller, until it reaches zero. At this point, further increases in x will lower total profit. So, the rule for finding the highest profit is to choose x such that the first derivative is zero.

The previous paragraph exemplifies a general and very important result for economics, known in mathematics as *Fermat's Theorem*. For a strictly concave function, $\tilde{f}(x)$, the value of x that maximizes $\tilde{f}(x)$, satisfies the *first order condition*, $\tilde{f}'(x) = 0$. In the profit function example above, we can find the profit maximizing value of x explicitly by solving,

$$\tilde{f}'(x) = aAx^{a-1} - p = 0, \text{for } x \text{ to get } x = (aA/p)^{1/(1-\mathrm{a})}.$$

EXAMPLES FROM THE TEXT

In Sect. 2.2 from Chap. 2, we can use the budget constraint to write the life-time utility of the household as a function of a single unconstrained choice variable, c_{1t},

$$U = \tilde{f}(c_{1t}) = \ln(c_{1t}) + \beta \ln(w_t - c_{1t}) + \beta \ln(R_t).$$

The first and second derivatives taken with respect to c_{1t} are

$$\tilde{f}'(c_{1t}) = -\frac{1}{c_{1t}} - \beta\frac{1}{w_t - c_{1t}}$$

$$\tilde{f}''(c_{1t}) = -\frac{1}{c_{1t}^2} - \beta\frac{1}{(w_t - c_{1t})^2} < 0.$$

Solving the first order condition for c_{1t}, $\tilde{f}'(c_{1t}) = 0$, gives the utility maximizing choice,

$c_{1t} = \frac{w_t}{1+\beta}$. Substituting the optimal choice of c_{1t} into the first and second period household budget constraints allows one to find the optimal choice of saving and second period consumption.

Multiple Choice Variables

Often economic agents are modelled as attempting to "do the best they can," more formally as maximizing some objective function, by choosing more than one variable. The basic approach when there is more than one choice variable is analogous to the one variable case. We illustrate the approach in the situation where there are two choice variables. In this case, the net benefit function has two arguments, x_1 and x_2, and is written as $\tilde{f}(x_1, x_2)$. The derivative of $\tilde{f}(x_1, x_2)$ with respect to *each* choice variable can be taken one at a time. These types of derivatives are called *partial derivatives*—they give the change in the function due to a change in one of the arguments, *holding all other arguments constant*.

One way of reinforcing the notion and the mechanics of taking a partial derivative is to think of a function with a *single argument* created from $\tilde{f}(x_1, x_2)$. This is done by holding x_2 constant. When x_2 is fixed at a certain value, it simply becomes a constant part of the newly defined function. For example, if we think of x_2 as fixed at the value $\bar{x}_2$, we can define the new function $h(x_1) \equiv \tilde{f}\left(x_1, \bar{x}_2\right)$. The partial derivative of $\tilde{f}(x_1, x_2)$ with respect to x_1 is then defined as $\tilde{f}_{x_1} \equiv h'(x_1)$ or, using a different notation, as $\dfrac{\partial \tilde{f}}{\partial x_1} \equiv h'(x_1)$. The second notation is a bit clumsy, but it is clearer in dynamic models where subscripts are used to denote time periods. Both types of notation are frequently used. Of course, the same procedure can be used to define the partial derivative with respect to x_2.

The partial derivatives are themselves typically functions of x_1 and x_2 and so they can be differentiated to get the *second partial derivatives*. There is a way of checking for the concavity of $\tilde{f}(x_1, x_2)$ that involves the second partial derivatives. This check is a bit complicated, so you need to trust that when we do maximization problems in the text, that we are using concave functions. However, if you build your own original models, you need to research the different ways of checking for concavity of functions with multiple choice variables.

If you are sure that $\tilde{f}(x_1, x_2)$ is a strictly concave function of x_1 and x_2, then you can identify the maximizing choices of x_1 and x_2 using the first order conditions in a manner perfectly analogous to the case with a function of just one variable. The first order conditions simply set the partial derivatives equal to zero,

$$\frac{\partial \tilde{f}}{\partial x_1} = 0 \text{ and } \frac{\partial \tilde{f}}{\partial x_2} = 0.$$

EXAMPLES FROM THE TEXT

In Chap. 2, the Cobb-Douglas production function is introduced,

$$Y_t = AK_t^{\alpha} M_t^{1-\alpha},$$

where Y denotes output, K denotes the capital stock rented, M denotes the hours of work hired, and where $A > 0$ and $0 < \alpha < 1$ are technological parameters.

The marginal product of an input is the increase in output that results from an increase in the use of an input. Formally, it is the partial derivative of the production function with respect to a particular input, holding other inputs constant. For a Cobb-Douglas production function, the marginal product of labor and the marginal product of capital are $\frac{\partial Y_t}{\partial M_t} = (1-\alpha)AK_t^{\alpha}M_t^{-\alpha}$ and $\frac{\partial Y_t}{\partial K_t} = \alpha AK_t^{\alpha-1}M_t^{1-\alpha}$ (see the rules for differentiating power functions given above). These expressions can be simplified somewhat by using algebra to write them in terms of the *capital intensity*, $k_t \equiv K_t/M_t$. The simplified expressions for the marginal products are, $\frac{\partial Y_t}{\partial M_t} = (1-\alpha)Ak_t^{\alpha}$ and $\frac{\partial Y_t}{\partial K_t} = \alpha Ak_t^{\alpha-1}$ (see the algebra rules for manipulating expressions with exponents given above).

We assume that markets are perfectly competitive in our production economy. As discussed in elementary economics, the notion of competitive markets applies not only to the markets for goods but also to the factor markets for labor and capital. The competitive assumption applied to the factor markets means that firms demand inputs to maximize profits taking as given the market prices of the inputs: the wage rate paid to labor (w) and rental rate on physical capital (r). No single firm is large enough to be able to influence market prices when they unilaterally change their production or input levels. The price of the economy's single output good is taken to be one. So we can think of output and revenue as being the same.

Given the competitive assumptions, the profit function can then be written as $Y_t - w_tM_t - r_tK_t$. Just as in the one-variable case, maximizing profits requires that firms hire capital and labor as long as the marginal benefit (marginal product) exceeds the marginal cost (factor price). Formally, the necessary first order conditions for profit maximization are

$$\alpha Ak_t^{\alpha-1} = r_t \text{ and } (1-\alpha)Ak_t^{\alpha} = w_t.$$

Constrained Maximization with Multiple Choice Variables

Let's extend the discussion from the previous section to the case where $f(x_1, x_2)$ is a strictly concave function of x_1 and x_2, but where the choice variables have to satisfy a resource constraints of the general form $F(x_1, x_2) = E$, where E is a positive constant. When resource constraints are present, there is a very important method that generates the first order conditions for the maximizing values of x_1 and x_2. It is called the *Lagrangian Method* , named after its inventor, the mathematician Joseph- Louis Lagrange. He showed that the first order conditions that must be satisfied by the maximizing values of x_1 and x_2 are

$$\frac{\partial f}{\partial x_1} = \lambda \frac{\partial F}{\partial x_1}, \quad \frac{\partial f}{\partial x_2} = \lambda \frac{\partial F}{\partial x_2}, \quad \text{and} \quad F(x_1, x_2) = E,$$

where λ is a variable called the *Lagrange multiplier.*

The first order conditions are easy to remember because they can reproduced by maximizing the *Lagrangian function*, $L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda[E - F(x_1, x_2)]$ with respect to x_1, x_2, λ. In other words, treat L as any other function and find the maximizing values by setting the partial derivatives of L to zero,

$$\frac{\partial L}{\partial x_1} = 0, \quad \frac{\partial L}{\partial x_2} = 0, \quad \text{and} \quad \frac{\partial L}{\partial \lambda} = 0.$$

These three equations, when written out and rearranged algebraically, are exactly the three first order conditions stated above.

EXAMPLES FROM THE TEXT

In Sect. 2.3 from Chap. 2, households maximize their lifetime utility by choosing the optimal consumption path over their two periods of life subject to their lifetime budget constraint. Matching the household's problem with the general set-up above we have

$$f(x_1, x_2) \equiv \ln(c_{1t}) + \beta \ln(c_{2t+1}), \quad F(x_1, x_2) \quad = \quad c_{1t} + \frac{c_{2t}}{R_t}, \quad \text{and} \quad E \equiv w_t$$

The Lagrangian function in our application is

$$L(c_{1t}, c_{2t+1}, \lambda_t) \quad = \quad \ln(c_{1t}) + \beta \ln(c_{2t+1}) \quad + \quad \lambda_t \left(w_t - c_{1t} - \frac{c_{2t+1}}{R_t} \right).$$

Differentiating and setting the partial derivatives equal to zero, gives us

$$\frac{1}{c_{1t}} = \lambda_t, \quad \frac{\beta}{c_{2t+1}} = \frac{\lambda_t}{R_t}, \quad \text{and} \quad c_{1t} + \frac{c_{2t+1}}{R_t} \quad = \quad w_t,$$

(see the rules for differentiating the natural log function given above). Solving these three equations for the three unknowns (c_{1t}, c_{2t+1}, λ_t), yields the optimal consumption demand functions and a value for the Lagrange multiplier, $c_{1t} = \frac{w_t}{1+\beta}, c_{2t+1} = \frac{\beta R_t w_t}{1+\beta}, \quad \lambda_t = \frac{1+\beta}{w_t}$.

A.3. Nonnegativity Constraints and Corner Solutions

The choice variables of economic agents are often restricted to be nonnegative values. The optimization approach taken in Sect. A.2 does not explicitly acknowledge this type of constraint on the choice variables. In many situations this is not a problem because, given the choice variables and the particular functions chosen, the optimal solutions naturally come out to be positive values. However, in some applications it is quite possible that some of the unconstrained optimal choice variables may take on negative values. This is not the proper solution if there are economic constraints preventing that possibility.

Fortunately, the Lagrangian method can be modified to account for nonnegativity constraints. The first order conditions with nonnegatvity constraints on x_1 and x_2 are

(i)

$$\frac{\partial L}{\partial x_1} \leq 0, x_1 \geq 0,$$

(ii)

$$\frac{\partial L}{\partial x_2} \leq 0, x_2 \geq 0,$$

and

(iii)

$$\frac{\partial L}{\partial \lambda} = 0.$$

where in (i) and (ii), *at least one* of the inequalities must be a strict equality. In the situation where the optimal values of both choices variables is strictly positive, then $x_1 > 0$ and $x_2 > 0$, so by the rule just stated $\frac{\partial L}{\partial x_1} = 0$ and $\frac{\partial L}{\partial x_2} = 0$, exactly as in the case where nonngegativity constraints are not accounted for. However, if an unconstrained choice of, say x_1, turns out to be negative, then the nonnegativity constraint binds and we have

$$\frac{\partial L}{\partial x_1} < 0, x_1 = 0.$$

This condition can be interpreted intuitively in the following way. Begin by thinking of $\frac{\partial L}{\partial x_1}$ as the marginal net benefit of increasing the value of x_1 (note that the Lagranian function incorporates both benefits and costs). If at $x_1 = 0, \frac{\partial L}{\partial x_1} > 0$, then the marginal benefit is positive and it is rational to increase x_1 above zero. However, if $\frac{\partial L}{\partial x_1} < 0$, then it is rational to *reduce* x_1 below zero in order to cause the total net benefit to rise. If this is not permitted, then the best the decision maker can do is set $x_1 = 0$. Because $x_1 = 0$ is at the end or at the "corner" of the permissible choices for x_1, this is referred as a *corner solution*.

EXAMPLES FROM THE TEXT

The approach used to handle choice variables that cannot be negative also works when a choice variable is constrained to exceed any particular value. In Chap. 4 we encounter a situation where schooling has a positive lower bound of $\bar{e}$, where $e_t \geq \bar{e}$. The rules for finding the optimal choice of e_t are

$$\frac{\partial L}{\partial e_t} \leq 0, e_t \geq \bar{e}.$$

A corner solution results unless the derivative (marginal net benefit) is positive when evaluated at the point $e_t = \bar{e}$. If $\frac{\partial L}{\partial e_t} > 0$ at $e_t = \bar{e}$, then the optimal choice is the interior solution $e_t = \frac{\theta\left(\eta\left(e_{t-1}/\bar{e}\right)^{\theta} - \gamma T\right)}{\gamma(1-\theta)} > \bar{e}$, the value for e_t that satisfies $\frac{\partial L}{\partial e_t} = 0$. If $\frac{\partial L}{\partial e_t} \leq 0$ at $e_t = \bar{e}$, then the best the household can do is choose the corner solution $e_t = \bar{e}$.

A.4. Total Differentials and Linear Approximations

If $y = f(x_1, x_2)$ is a differentiable function of x_1 and x_2, one can define the *total differential* of f as

$$dy = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2,$$

where dy, dx_1, and dx_2 are real variables that are interpreted as "changes" in the original variables. The concept of the total differential extends naturally to the case where the function has many arguments or independent variables.

If one imagines that the total differential is taken at a particular point where $x_1 = \bar{x}_1$ and $x_2 = \bar{x}_2$, then it can be related to the notion of a *linear approximation* of $f(x_1, x_2)$,

$$y = f\left(\bar{x}_1, \bar{x}_2\right) + \frac{\partial f}{\partial x_1}\left(\bar{x}_1, \bar{x}_2\right) dx_1 + \frac{\partial f}{\partial x_2}\left(\bar{x}_1, \bar{x}_2\right) dx_2,$$

where dx_1 and dx_2 are interpreted as deviations from the values $x_1 = \bar{x}_1$ and $x_2 = \bar{x}_2$, and the partial derivatives are evaluated at the point $(\bar{x}_1, \bar{x}_2)$. Note that, analogous to the interpretations of dx_1 and dx_2, it is natural to think of dy as $y - f\left(\bar{x}_1, \bar{x}_2\right)$.

EXAMPLES FROM THE TEXT

In Sect. 6.3 from Chap. 6, we analyze a system of nonlinear difference equations where we cannot explicitly solve for future values of the state variables in terms of current values. In this situation one can conduct a qualitative analysis of the system in the neighborhood of a steady state by taking a linear approximation to the nonlinear system of equations.

To see how this works in a simpler setting, consider taking a linear approximation to the nonlinear difference equation from Chap. 2,

$$k_{t+1} = f(k_t) \equiv Bk_t^{\alpha}.$$

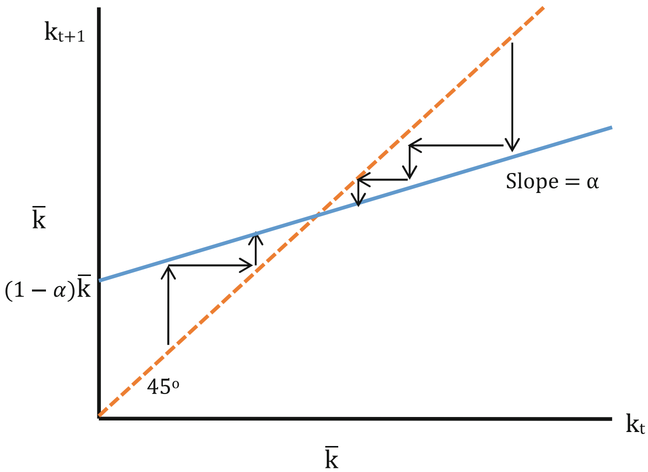
Taking a linear approximation of f in the neighborhood of the steady state gives us

$$k_{t+1} \approx f(\bar{k}) + f'(\bar{k})(k_t - \bar{k}) = \bar{k} + \alpha B\bar{k}^{\alpha-1}(k_t - \bar{k}).$$

We know the steady state k is $\bar{k} = B^{\frac{1}{1-\alpha}}$, implying that $\alpha B\bar{k}^{\alpha-1} = \alpha BB^{-1} = \alpha$. The linear difference equation that serves to approximate the behavior of the nonlinear difference equation near the steady state can then be written as

$$k_{t+1} = (1-\alpha)\bar{k} + \alpha k_t.$$

The difference equation is sketched below. Note that it exhibits the stability property possessed by the original nonlinear difference equation (near the steady state).



A.5. L'Hospital's Rule

On occasion one encounters a ratio of functions or expressions that take on an *indeterminate form* at a point of interest. An indeterminate form is one where the ratio becomes $\frac{0}{0}$ or $\frac{\infty}{\infty}$. In some cases indeterminate forms actually do have a determinate value that is simply not immediately obvious. *L'Hospital's Rule* indicates when this might be true.

The rule says that if you have two differentiable expressions, $f(x)$ and $h(x)$, and at a particular value of x, say $x = x_0$, the ratio $\frac{f(x)}{h(x)}$ takes an indeterminate form, then $\lim_{x \to x_0} \frac{f(x)}{h(x)} = \lim_{x \to x_0} \frac{f'(x)}{h'(x)}$. The result is useful because sometimes the ratio of derivatives has a determinate form.

EXAMPLES FROM THE TEXT

In Sect. 2.6 and *Problem* 7 of Chap. 2, we introduced a more general lifetime utility function with a single period utility flow from consumption of the form,

$$u_t = \frac{\left(c_t^{1-1/\sigma} - 1\right)}{(1 - 1/\sigma)}.$$

The motivation for needing a more general utility function is provided in the text, but part of the reason for its unusual form is to allow the logarithmic utility function, that we use in most of our models, to appear as a special case. Using L'Hospital Rule one can show that $u_t = \ln c_t$, when $\sigma = 1$.

To see this, first note that when $\sigma = 1$, the utility function has the indeterminate form $\frac{0}{0}$. Second, we need to use the result that the exponential function and the natural log functions are inverses of each other, i.e. $x^a = e^{a \ln x}$. This means we can write $c_t^{1-1/\sigma}$ as $e^{(1-1/\sigma)\ln c_t}$. Third, the rule for differentiating the exponential function $f(x) = e^{ax}$, is $f'(x) = ae^{ax}$. Finally, to apply the result, think of the expressions in the numerator and the denominator as functions of σ.

Now, we can write utility as

$$u_t = \frac{\left(e^{(1-1/\sigma)\ln c_t} - 1\right)}{(1 - 1/\sigma)}.$$

Differentiating the numerator and the denominator with respect to σ and then taking the ratio of the two derivatives gives

$$\frac{\frac{1}{\sigma^2}\ln c_t e^{(1-1/\sigma)\ln c_t}}{\frac{1}{\sigma^2}} = \ln c_t e^{(1-1/\sigma)\ln c_t}.$$

At $\sigma = 1$, the ratio is $u_t = \ln c_t$, because $e^0 = 1$.

A.6. Quadratic Equations

Some equations in the unknown variable x can be written in the following quadratic form

$$ax^2 + bx + c = 0,$$

where $a \neq 0$. Mathematically, there are two solutions for x that satisfy the equation, although one or both may not make sense as solutions to an economic problem. The mathematical solution are given by the *quadratic formula*,

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

EXAMPLES FROM THE TEXT

In *Problem* 19 of Chap. 2 we consider a special case of the CES utility function where $\sigma = \frac{2-\alpha}{1-\alpha}$. With $\delta = 1$ and $\sigma = \frac{2-\alpha}{1-\alpha}$, we can write the transition equation (9) from Chap. 2 as $k_t = \frac{(1-\alpha)Ak_{t-1}^{\alpha}}{n(1+d)} \frac{1}{1+\beta^{-\sigma}(\alpha A)^{1-\sigma}k_t}$. This expression can be written as $\beta^{-\sigma}(\alpha A)^{1-\sigma}k_t^2 + k_t + \frac{-(1-\alpha)Ak_{t-1}^{\alpha}}{n(1+d)} = 0$, which is a quadratic equation in the unknown k_t. Applying the quadratic formula reveals that there is only one positive solution, $k_t = \frac{\left(1 + 4\frac{\beta^{-\sigma}(\alpha A)^{1-\sigma}(1-\alpha)Ak_{t-1}^{\alpha}}{n(1+d)}\right)^{1/2} - 1}{2\beta^{-\sigma}(\alpha A)^{1-\sigma}}$.

A.7. Infinite Series

A *sequence* is an ordered list of terms, $a_0, a_1, a_2, \cdot, \cdot, \cdot, a_n$. A special case of a sequence is one where consecutive terms have the same ratio, known as a *geometric sequence*. This is possible when the terms of the sequence have a common base value that is raised to an increasing power as follows:

$a_0 = a^0 = 1$, $a_1 = a^1 = a$, $a_2 = a^2$, $a_3 = a^3, \cdot, \cdot, \cdot, a_n = a^n$. So the ratio of consecutive terms is always a.

Of more direct interest to us is the *sum* of a geometric sequence known as a *geometric series*, defined as

$$S_n = \sum_{i=0}^{n} a^i = 1 + a + \cdots + a^n.$$

Note that $S_n - aS_n = 1 - a^{n+1}$, so

$$S_n = \frac{1 - a^{n+1}}{1 - a}.$$

Finally, note when $0 \leq a < 1$, then if $n \to \infty$, the *infinite geometric series* is

$$S_\infty = \frac{1}{1-a}.$$

EXAMPLES FROM THE TEXT

In Eq. (2.31) from Sect. 2.6 of Chap. 2, we encounter an infinite series of the form $\left\{1 + \left(\frac{n}{R}\right)\left[\frac{\beta R}{n}\right]^{\sigma} + \left(\frac{n}{R}\right)^2\left[\frac{\beta R}{n}\right]^{2\sigma} + \cdots\right\}$ in an equation that can be used to solve for the consumption of the first generation of a dynastic chain linked by intergenerational altruism,

$$\Psi_1^{-1} c_{1t}\left\{1 + \left(\frac{n}{R}\right)\left[\frac{\beta R}{n}\right]^{\sigma} + \left(\frac{n}{R}\right)^2\left[\frac{\beta R}{n}\right]^{2\sigma} + \cdots\right\} = W_\infty. \tag{2.31}$$

The geometric sum, in the curly brackets of (2.31), is finite provided $\beta^\sigma\left(\frac{n}{R}\right)^{1-\sigma} < 1$ or $\beta^{\sigma/(1-\sigma)} \leq \frac{R}{n}$. If $R > n$, then this condition holds when $\sigma \leq 1$ and $\beta \leq 1$ Under these conditions, the value of the infinite series is

$\frac{1}{1-\beta^\sigma\left(\frac{n}{R}\right)^{1-\sigma}}$ and the solution for consumption is $c_{1t} = \Psi_1\left(1 - \beta^\sigma\left(\frac{n}{R}\right)^{1-\sigma}\right)W_\infty$.

Index

S. Das et al., *Economic Growth and Development*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-319-89755-4