CHAPTER 21

Structural Transformations and Market Failures in Development

Together with the process of economic development and the changes in the structure of production, there is also a transformation of the economy, which both involves major social changes and induces greater (and perhaps more "complex") coordination of economic activities. Loosely speaking, we can think of a society that is relatively developed as functioning along (or at any rate, nearer) the frontier of its production possibilities set, while a less-developed economy may be in the interior of its "notional" production possibilities set. This may be because certain arrangements necessary for an economy to reach the frontier of its production possibility set require a large amount of capital or some specific technological advances (in which case, even though we may think of the society as functioning in the interior of its production possibility set, this may not be the outcome of market failure, thus the qualifier "notional" in the previous sentence). Alternatively, less developed economies may be in the interior of their production possibility set because these societies are subject to severe market failures. In this chapter, I discuss these approaches to economic development.

I first focus on various dimensions of structural transformations and how these may be limited by the amount of capital or technology available in a society. I will then discuss a number of approaches suggesting that less-developed economies might be suffering disproportionately from market failures or may even be "stuck" in "development traps". In this context, I will also discuss differences between models with multiple equilibria and with multiple steady states.

The topics covered in this chapter are part of a large and diverse literature. My purpose is not to do justice to this literature, but to emphasize how certain major structural transformations take place as part of the process of economic development and also highlight the potential importance of market failures in this process. Given this objective and the large number of potential models, my choice of models is selective and my treatment will be more informal than the rest of the book. In addition, I often make reduced-form assumptions in order to keep the exposition brief and simple.

21.1. Financial Development

An important aspect of the structural transformation brought about by economic development is a change in financial relations and deepening of financial markets. Section 17.6 in Chapter 17 already presented a model where economic growth goes hand-in-hand with financial deepening. However, the model in that section only focused on a specific aspect of the role of financial institutions. In general, financial development brings about a number of complementary changes in the economy. First, there is greater depth in the financial market, allowing better diversification of aggregate risks, a feature also emphasized in the model of Section 17.6. Second, one of the key roles of financial markets is to allow risk sharing and consumption smoothing for individuals. In line with this, financial development also allows better diversification of *idiosyncratic risks*. Section 17.6 showed that better diversification of aggregate risks leads to a better allocation of funds across sectors/projects. Similarly, better sharing of idiosyncratic risks leads to a better allocation of funds across individuals. Third, financial development might also reduce credit constraints on investors and thus may directly enable the transfer of funds to individuals with better investment opportunities. The second and the third channels not only affect the allocation of resources in the society but also the distribution of income, because diversification of idiosyncratic risks and relaxation of credit market constraints might lead to better income and risk sharing. On the other hand, as the possibility of such risk-sharing arrangements reduce consumption risk, individuals might take riskier actions, also potentially affecting the distribution of income.

To provide a brief introduction to these issues, I now present a simple model of financial development, focusing on the diversification of idiosyncratic risks and complementing the analysis in Section 17.6. The model is inspired by the work of Townsend (1979) and Greenwood and Jovanovic (1990). It will illustrate how financial development takes place endogenously and interacts with economic growth, and will also provide some simple insights about the implications of financial development for income distribution. Given the similarity of the model to that in Section 17.6, my treatment here will be relatively informal.

I consider an OLG economy in which each individual lives for two periods and has preferences given by

$$\mathbb{E}_t U_t(c(t), c(t+1)) = \log c(t) + \beta \mathbb{E}_t \log c(t+1), \tag{21.1}$$

where $c(t)$ denotes the consumption of the unique final good of the economy and $\mathbb{E}_t$ denotes the expectation operator given time t information.

There is no population growth and the total population of each generation is normalized to 1. Let us assume that each individual is born with some labor endowment l. The distribution of endowments across agents is given by the distribution function $G(l)$ over some support $[\underline{l}, \bar{l}]$ This distribution of labor endowments is constant over time with mean $L = 1$ and labor is supplied inelastically by all individuals in the first period of their lives. In the second period of their lives, individuals simply consume their capital income.

The aggregate production function of the economy is given by

$$Y(t) = K(t)^{\alpha} L(t)^{1-\alpha} = K(t)^{\alpha},$$

where $\alpha \in (0,1)$ and the second equality uses the fact that total labor supply will be equal to 1 at each date. As in Section 17.6, the only risk is in transforming savings into capital, thus the lifecycle of an individual looks identical to that shown in Figure 17.3 in that section. Moreover, suppose that agents can either save all of their labor earnings from the first period of their lives using a safe technology with rate of return q (in terms of capital at the next date) or invest all of their labor income in the risky technology with return $Q + \varepsilon$, where ε is a mean zero independently and identically distributed stochastic shock and as in Section 17.6, we assume that $Q > q$. This implies that the risky technology is more productive. The assumption that individuals have to choose one of these two technologies rather than dividing their savings between the two is made for simplicity (see Exercise 21.1).

Although the model looks very similar to that in Section 17.6, there is a crucial difference. Because ε is identically and independently distributed *across individuals*, if individuals could pool their resources, they could perfectly diversify idiosyncratic risks. In particular, if a large number (a continuum) of individuals pooled their resources, they would guarantee an average return of Q. Let us assume that this is not possible because of a standard *informational problem*—the actual return of an individual's saving decision is not observed by others unless some financial monitoring is undertaken. Let us assume that this type of financial monitoring costs $\xi > 0$ for each individual. This implies that by paying the cost of ξ, each individual can join the financial market (or in the language of Townsend, he can become part of a "financial coalition"). In this case, the actual returns of his savings become fully observable. Intuitively, this cost captures the fixed costs that individuals have to pay to be engaged in financial markets as well as the fixed costs associated with monitoring or being monitored. An immediate implication of this specification is that joining the financial markets is more attractive for richer individuals, since the fixed cost is less important for them. This feature is both plausible and also generates predictions consistent with microdata, where we observe richer individuals investing in more complex financial securities.

If an individual does not join the financial markets, then no other agent in the economy can observe the realization of the returns on his savings. In this case, no financial contract for sharing of idiosyncratic risks is possible, since such a contract would involve agents that have a high (realized) value of ε making transfers to those who are unlucky and have low realized values of ε. However, without monitoring, each agent will claim to have a low value of ε, thus receive ex post payments. The anticipation of this type of opportunistic behavior prevents any risk sharing in the absence of monitoring.

Let us also assume that ε has a distribution with positive probability of $\varepsilon = -Q$, so that if an individual undertakes the risky investments, there is a positive probability that all his savings will be lost. This implies that without some type of risk sharing, individuals will choose the safe project. This observation simplifies the analysis of the model. Suppose that the economy starts with some initial capital stock of $K(0) > 0$, so an individual with labor endowment l_i will have labor earnings of $W_i(0) = w(0)\, l_i$, where

$$w(t) = (1-\alpha) K(t)^{\alpha} \tag{21.2}$$

is the competitive wage rate at time t. After labor incomes are realized, individuals first make their savings decisions and then choose which assets to invest in. The preferences in (21.1) imply that individuals will save a constant fraction $\beta/(1+\beta)$ of their income regardless of their income level or the rate of return (in particular, independent of whether they are investing in the risky or the safe asset). In view of this, the value to not participating in the financial markets for individual i at time t is

$$V_i^N(W_i(t), R(t+1)) = \log\left(\frac{1}{1+\beta}W_i(t)\right) + \beta\log\left(\frac{\beta R(t+1)\,q}{1+\beta}W_i(t)\right),$$

which takes into account that the rate of return on capital in the second period of the life of the individual will be $R(t+1)$ and the individual will receive a gross return q on his savings of $\beta W_i(t)/(1+\beta)$. Next, suppose that there are sufficiently many (that is, a positive measure of) other individuals taking part in financial markets. Then, when the individual decides to take part in financial markets, his value will be

$$V_i^F(W_i(t), R(t+1)) = \log\left(\frac{1}{1+\beta}(W_i(t)-\xi)\right) + \beta\log\left(\frac{\beta R(t+1)\,Q}{1+\beta}(W_i(t)-\xi)\right),$$

which takes into account that the individual will have to spend the amount ξ out of his labor income on the cost of joining the financial market, leaving him a net income of $W_i(t)-\xi$. He will then save a fraction $\beta/(1+\beta)$ of this, but in return, he will receive the higher return Q for sure. The reason why the individual receives Q, rather than a risky return, is because, conditional on joining the financial market, each individual is able to fully diversify his idiosyncratic risks. The comparison of these two expressions gives the threshold level

$$W^* \equiv \frac{\xi}{1-(q/Q)^{\beta/(1+\beta)}} > 0, \tag{21.3}$$

such that individuals with first-period earnings greater than W^* will join the financial market and those with less than W^* will not. A notable feature of this threshold W^* is that it is independent of the rate of return on capital in the second period of the lives of the individuals, R. This is an implication of log preferences in (21.1).

Given the behavior of individuals concerning whether they will join the financial market, let us next determine the evolution of the economy by studying the evolution of individual earnings. Individual earnings are determined by two factors: labor endowments and the capital stock at time t, which gives the wage per unit of labor, $w(t)$, as in (21.2). Given $w(t)$, the fraction of individuals who will join the financial market at time t, $g^F(t)$, is given by the fraction of individuals who have $l_i \geq W^*/w(t)$. Alternatively, using the fact that labor endowments have a distribution given by $G(\cdot)$, the fraction of individuals investing in financial markets is obtained as

$$g^F(t) \equiv 1 - G\left(\frac{W^*}{w(t)}\right) = 1 - G\left(\frac{W^*}{(1-\alpha)K(t)^{\alpha}}\right). \tag{21.4}$$

In view of this and defining $\chi(t) \equiv W^{*}/(1-\alpha)K(t)^{\alpha}$, the capital stock at time $t+1$ can be written as

$$K(t+1) = \frac{\beta}{1+\beta}\left[q\int_{\underline{l}}^{\chi(t)} l dG(l) + Q\int_{\chi(t)}^{\bar{l}} (l-\xi)\, dG(l)\right](1-\alpha)K(t)^{\alpha}, \tag{21.5}$$

which takes into account that all individuals with labor endowment less than $\chi(t)$ choose the safe project and receive the gross return q on their savings, while those above this threshold spend ξ on monitoring and receive the higher return Q. It can be verified that $K(t+1)$ is increasing in $K(t)$ and there will be growth in the capital stock (and thus output) of the economy provided that $K(t)$ is less than the "steady-state" level of capital; see Exercise 21.2).

Now inspection of the accumulation equation (21.5) together with the threshold rule for joining the financial market leads to a number of interesting conclusions.

1. As $K(t)$ increases, that is, as the economy develops, equation (21.4) implies that more individuals will join the financial market. Consequently, a greater level of capital will lead to more risk taking, but these risks will also be shared better. More importantly, economic development also induces a better composition of investment as a greater fraction of the individuals start using their savings more efficiently. Thus with a mechanism similar to that in Section 17.6, economic development improves the allocation of funds in the economy and increases productivity. Consequently, this model, like the one in Section 17.6, implies that economic development and financial development go hand-in-hand.

2. However, there is also a distinct sense in which the economy here allows for a potential causal effect of financial development on economic growth. Imagine that societies differ according to their ξ's, which can be interpreted as a measure of the institutionally- or technologically-determined costs of monitoring or some other costs associated with financial transactions that may depend on the degree of investor protection. Societies with lower ξ's will have a greater participation in financial markets and this will endogenously increase their productivity. Thus, while the equilibrium behavior of financial and economic development are jointly determined, differences in financial development driven by exogenous institutional factors related to ξ will have a potential causal effect on economic growth.

3. As noted above, at any given point in time it will be the richer agents—those with greater labor endowment—that will join the financial market. Therefore, initially, the financial market will help those who are already well-off to increase the rate of return on their savings. This can be thought of as the *unequalizing* effect of the financial market.

4. The fact that participation in financial markets increases with $K(t)$ also implies that as the economy grows, at least at the early stages of economic development, the unequalizing effect of financial intermediation will become stronger. Therefore, presuming that the economy starts with relatively few rich individuals, the first expansion of the financial market will increase the level of overall inequality in the economy as a greater fraction of the agents in the economy now enjoy the greater returns.

5. As $K(t)$ increases even further, eventually the *equalizing* effect of the financial market starts operating. At this point, the fraction of the population joining the financial market and enjoying the greater returns is steadily increasing. If the steady-state level of capital stock K^* is such that $\underline{l} \geq W^*/(1-\alpha)(K^*)^\alpha$, then eventually all individuals will join the financial market and receive the same rate of return on their savings.

The last two observations are interesting in part because the relationship between growth and inequality is a topic of great interest to development economics (one to which I return later in this chapter). One of the most important ideas in this context is that of the *Kuznets curve*, which claims that economic growth first increases and then reduces income inequality in the society. Whether the Kuznets curve is a good description of the relationship between growth and inequality is a topic of current debate. While many European societies seem to have gone through a phase of increasing and then decreasing inequality during the 19th century, the evidence for the 20th century is more mixed. The last two observations show that a model with endogenous financial development based on risk sharing among individuals can generate a pattern consistent with the Kuznets curve. Whether there is indeed a Kuznets curve in general, and if so, whether the mechanism highlighted here plays an important role in generating this pattern are areas for future theoretical and empirical work.

21.2. Fertility, Mortality and the Demographic Transition

Chapter 1 highlighted the major questions related to growth of income per capita over time and its dispersion across countries today. Our focus so far has been on these per capita income differences. Equally striking differences exist in the level of population across countries and over time. Figure 21.1 uses data from Maddison (2002) and shows the levels and the evolution of population in different parts of the world over the past 2000 years. The figure is in log scale, so a linear curve indicates a constant rate of population growth. The figure shows that starting about 250 years ago there is a significant increase in the population growth rate in many areas of the world. This more rapid population growth continues in much of the world, but importantly, the rate of population growth slows down in Western Europe sometime in the 19th century (though, thanks partly to immigration, not so in the Western Offshoots). There is no similar slow down of population growth in less-developed parts of the world. On the contrary, in many less developed nations, the rate of population growth seems to have increased over the past 50 years or so. We have already discussed one of the reasons for this in Chapter 4—the spread of antibiotics, basic sanitation and other health-care measures around the world that reduced the very high mortality rates in many countries. However, equally notable is the *demographic transition*, which, in the course of the 19th century, reduced fertility in Western Europe. Why population has grown slowly and then accelerated to reach a breakneck speed of growth over the past 150 years and why population growth rates differ across countries are major questions for economic development.

In this section, I present the most basic approaches to population dynamics and fertility. I first discuss a simple version of the famous Malthusian model and then use a variant of this

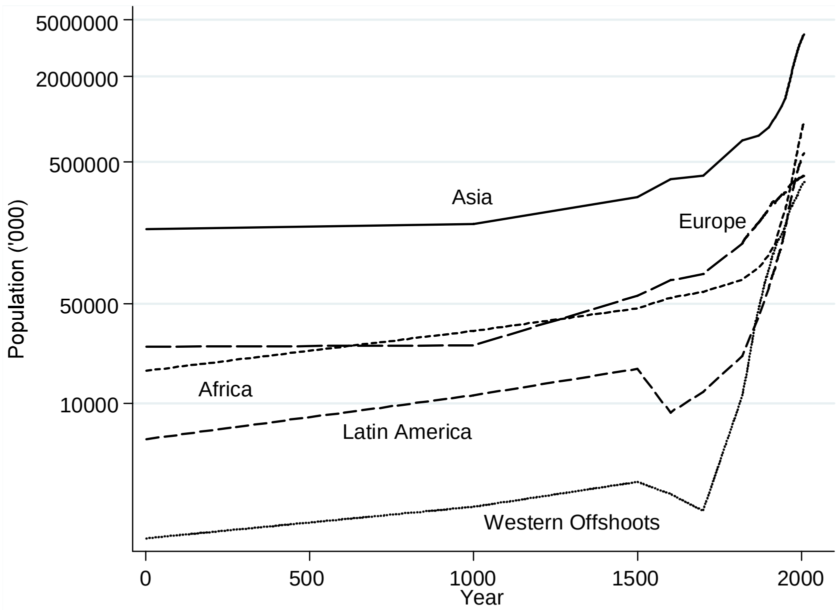

Figure 21.1. Total population in different parts of the world over the past 2000 years.

model to investigate potential causes of the demographic transition. Thomas Malthus was one of the most brilliant and influential economists of the 19th century and is responsible for one of the first general equilibrium growth models. The next subsection will present a version of this model. The Malthusian model is responsible for earning the discipline of economics the name "the dismal science" because of its dire prediction that population will adjust up or down (by births or deaths) until all individuals are at the subsistence level of consumption. Nevertheless, this dire prediction is not the most important part of the Malthusian model. Instead, at the heart of this model is the *negative relationship* between income per capita and population, which is itself endogenously determined. In this sense, it is closely related to the Solow and the neoclassical growth models, augmented with a behavioral rule that determines the rate of population growth. It is this less extreme version of the Malthusian model that will be presented in the next subsection. I then enrich this model by the important and influential idea due to Gary Becker that there is a tradeoff between the quantity and quality of children and that this tradeoff changes over the process of development. I show how a simple model can incorporate the notion that over the course of development parents may start valuing the quality (human capital) of their offspring more and how this may lead to a pattern reminiscent to the demographic transition.

21.2.1. A Simple Malthusian Model. Consider the following non-overlapping generations model that starts with a population of $L(0) > 0$ at time $t = 0$. A representative individual living at time t supplies one unit of labor inelastically and has utility

$$c(t)^{\beta}\left[y(t+1)\,n(t+1)-\frac{1}{2}\eta_0 n(t+1)^2\right], \tag{21.6}$$

where $c(t)$ denotes the consumption of the unique final good of the economy by the individual himself, $n(t+1)$ denotes the number of offspring the individual begets and $y(t+1)$ is the income of each offspring, and $\beta > 0$ and $\eta_0 > 0$. The last term in square brackets is the child-rearing costs and is assumed to be convex to reflect the fact that the costs of having more and more children will be higher (for example, because of time constraints of parents, though one can also make arguments for why the costs of child-rearing might exhibit increasing returns to scale over a certain range). Clearly, these preferences introduce a number of simplifying assumptions. First, each individual is allowed to have as many offspring as he likes, which is unrealistic because it does not restrict the number of offspring to a natural number. The technology also does not incorporate possible specialization in child-rearing and market work within the family. Second, these preferences introduce the "warm glow" type altruism we encountered in Chapter 9, so that parents receive utility not from the future utility of their offspring, but from some characteristic of their offspring. Here it is a transform of the total income of all the offspring that features in the utility function of the parent. Third, the costs of child-rearing are in terms of "utils" rather than forgone income, and current consumption multiplies both the benefits and the costs of having additional children. This feature, which is motivated by balanced growth type reasoning, implies that the demand for children will be independent of current income (otherwise, growth will automatically lead to greater demand for children). All three of these assumptions are adopted for simplicity. I have also written the number of offspring that an individual has a time t as $n(t+1)$, since this will determine population at time $t+1$.

Each individual has one unit of labor and there are no savings. The production function for the unique good takes the form

$$Y(t) = Z^{\alpha} L(t)^{1-\alpha}, \tag{21.7}$$

where Z is the total amount of land available for production and $L(t)$ is total labor supply. There is no capital and land is introduced in order to create diminishing returns to labor, which is an important element of the Malthusian model. Without loss of generality, I normalize the total amount of land to $Z = 1$. A key question in models of this sort is what happens to the returns to land. The most satisfactory way of dealing with this problem would be to allocate the property rights to land among the individuals and let them bequeath this to their offspring. This, however, introduces another layer of complication, and since my purpose here is to illustrate the basic ideas, I follow the unsatisfactory assumption often made in the literature, that land is owned by another set of agents, whose behavior will not be analyzed here.

By definition, population at time $t+1$ is given as

$$L(t+1) = n(t+1)\,L(t), \tag{21.8}$$

which takes into account new births as well as the death of the parent.

Labor markets are competitive, so the wage at time $t+1$ is given as

$$w(t+1) = (1-\alpha) L(t+1)^{-\alpha}. \tag{21.9}$$

Since there is no other source of income, this is also equal to the income of each individual living at time $t+1$, $y(t+1)$. Thus an individual with income $w(t)$ at time t will solve the problem of maximizing (21.6) subject to the constraint that $c(t) \leq w(t)$, together with $y(t+1) = (1-\alpha) L(t+1)^{-\alpha}$. Naturally, in equilibrium $n(t+1)$ must be consistent with $L(t+1)$ according to equation (21.8). Individual maximization implies

$$n(t+1) = (1-\alpha)\eta_0^{-1} L(t+1)^{-\alpha}.$$

Now substituting for (21.8) and rearranging, we obtain

$$L(t+1) = (1-\alpha)^{\frac{1}{1+\alpha}} \eta_0^{-\frac{1}{1+\alpha}} L(t)^{\frac{1}{1+\alpha}}. \tag{21.10}$$

This equation implies that $L(t+1)$ is an increasing concave function of $L(t)$. In fact, the law of motion for population implied by (21.10) resembles the dynamics of capital-labor ratio in the Solow growth model (or the OLG model) and is plotted in Figure 21.2. The figure makes it clear that starting with any $L(0) > 0$, there exists a unique globally stable state state L^* given by

$$L^* \equiv (1-\alpha)^{1/\alpha} \eta_0^{-1/\alpha}. \tag{21.11}$$

If the economy starts with $L(0) < L^*$, then population will slowly (and monotonically) adjust towards this steady-state level. Moreover, (21.9) shows that as population increases wages fall. If in contrast, $L(0) > L^*$, then the society experiences a decline in population and rising real wages. It is straightforward to introduce shocks to population and show that in this case, the economy will fluctuate around the steady-state population level L^* (with an invariant distribution depending on the distribution of the shocks) and experience cycles reminiscent to the Malthusian cycles, with periods of increasing population and decreasing wages followed by periods of decreasing population and increasing wages (see Exercise 21.3).

The main difference of this model from the simplest (or the crudest) version of the Malthusian model is that there is no biologically determined subsistence level of consumption. The steady-state level of consumption instead reflects technology and preferences, and is given by

$$c^* = (1-\alpha)(L^*)^{-\alpha} = \eta_0.$$

21.2.2. The Demographic Transition. To study the demographic transition, I now introduce a quality-quantity tradeoff along the lines of the ideas suggested by Gary Becker. Each parent can choose his offspring to be unskilled or skilled. To make them skilled, the parent has to exert the additional effort for child-rearing denoted by $e(t) \in \{0,1\}$. If he chooses not to do this, his offspring will be unskilled.

The total population of unskilled individuals at time t is denoted by $U(t)$ and the total population of the skilled is denoted by $S(t)$, clearly with

$$L(t) = U(t) + S(t).$$

Figure 21.2. Population dynamics in this simple Malthusian model.

The second modification is that there are now two production technologies that can be used for producing the final good. The Malthusian (traditional) technology is still given by (21.7) and any worker can be employed with this technology. The modern technology is

$$Y^M(t) = X(t)\,S(t)\,. \tag{21.12}$$

This equation implies that productivity in the modern technology is potentially time varying and also states that only skilled workers can be employed with this technology. It also imposes that all skilled workers will be employed with this technology.[1]

To model the quality-quantity tradeoff, individual preferences are now modified from equation (21.6) to

$$c(t)^{\beta}\left[y(t+1)(n(t+1)-1) - \frac{1}{2}(\eta_0(1-e(t)) + \eta_1 X(t+1)\,e(t))\,n(t+1)^2\right]. \tag{21.13}$$

This formulation of the preferences states that if the individual decides to invest in his offspring's skills, instead of the fixed cost η_0, he has to pay a cost that is proportional to the amount of knowledge $X(t+1)$ that the offspring has to absorb to use the modern technology. I assume that η_1 is sufficiently greater than η_0, and in particular, that $X(0)\,\eta_1 > \eta_0$, so that even at the initial level of the modern technology rearing a skilled child is more costly than an unskilled child.

[1]This need not be true in general, since wages in the traditional sector may be higher. However, in equilibrium this will never be the case because parents would not choose to exert the additional effort to endow their offspring with skills if they would then work in the traditional sector. To keep the exposition simple (and with a slight abuse of notation), equation (21.12) already incorporates the fact that, in equilibrium, all skilled workers will be employed in the modern sector.

Finally, I assume learning-by-doing is external as in Romer (1986a), so that

$$X(t+1) = \kappa S(t), \tag{21.14}$$

which implies that the improvement in the technology of the modern sector is a function of the number of skilled workers employed in this sector. This type of reduced-form assumption is clearly unsatisfactory, but as noted above, one could get similar results with an endogenous technology model with the market size effect. Another important feature of this production function is that it does not use land. This assumption is consistent with the fact that most modern production processes make little use of land, instead relying on technology, physical capital and human capital.

The output of the traditional and the modern sectors are perfect substitutes—they both produce the same final good. In view of the observation that all unskilled workers will work in the traditional sector and all skilled workers will work in the modern sector, wages of skilled and unskilled workers at time t are

$$w^U(t) = (1-\alpha)\,U(t)^{-\alpha}, \text{ and} \tag{21.15}$$

$$w^S(t) = X(t), \tag{21.16}$$

where (21.15) is identical to (21.9) in the previous subsection, except that it features only the unskilled workers instead of the entire labor force.

Let us next turn to the fertility and quality-quantity decisions of individuals. As before, current income has no effect on fertility and quality-quantity decisions. Thus we do not need to distinguish between high-skill and low-skill parents. Using this observation, let us simply look at the optimal number of offspring that an individual will have when he chooses $e(t) = 0$. This is given by

$$n^U(t+1) = w^U(t+1)\,\eta_0^{-1} = (1-\alpha)\,U(t+1)^{-\alpha}\,\eta_0^{-1}, \tag{21.17}$$

where the second line uses (21.15). Instead, if the parent decides to exert effort $e(t) = 1$ and invest in the skills of his offspring, then he will choose the number of offspring equal to

$$n^S(t+1) = w^S(t+1)\,X(t+1)^{-1}\,\eta_1^{-1} = \eta_1^{-1}. \tag{21.18}$$

The comparison of equations (21.17) and (21.18) suggests that unless unskilled wages are very low, an individual who decides to provide additional skills to his offspring will have fewer offspring. This is because bringing up skilled children is more expensive (i.e., because η_1 is sufficiently larger than η_0). Thus the comparison of these two equations captures the quality-quantity tradeoff.

Substituting these equations back into the utility function (21.13), we obtain the utility from the two strategies (normalized by consumption, that is, the utility divided by $c(t)^\beta$) as

$$V^U(t) = \frac{1}{2}(1-\alpha)^2\,U(t+1)^{-2\alpha}\,\eta_0^{-1} \text{ and } V^S(t) = \frac{1}{2}X(t+1)\,\eta_1^{-1}.$$

Inspection of these two expressions shows that in equilibrium, some workers must be unskilled, since otherwise V^U would become infinite. Therefore, in equilibrium

$$V^U(t) \geq V^S(t). \tag{21.19}$$

This equilibrium condition implies that there are two possible configurations. First, $X(0)$ can be so low that (21.19) will hold as a strict inequality. In this case, all individuals will be unskilled. The condition for this inequality to be strict is

$$X(0)\,\eta_1^{-1} < (1-\alpha)^2 L(1)^{-2\alpha}\,\eta_0^{-1},$$

which uses the fact that when there are no skilled workers, there is no production in the modern sector and thus $X(1) = X(0)$. If this inequality satisfied, there would be no skilled children at date $t = 0$. However, as long as $L(1)$ is less than L^* as given in (21.11), population will grow. It is therefore possible that at some point (21.19) holds with equality. The condition for this never to happen is that

$$X(0)\,\eta_1^{-1} < (1-\alpha)^2 (L^*)^{-2\alpha}\,\eta_0^{-1}. \tag{21.20}$$

In this case, the law of motion of population is identical to that in the previous subsection and there is never any investment in skills. We can think of this is a pure Malthusian economy.

If, on the other hand, condition (21.20) is not satisfied, then at least at some point individuals will start investing in the skills of their offspring. From then on, (21.19) must hold as equality. Let the fraction of parents having unskilled children at time t be denoted by $u(t+1)$. Then, by definition

$$\begin{aligned} U(t+1) &= u(t+1)\left(n^U(t+1)-1\right)L(t) \\ &= u(t+1)^{1/(1+\alpha)}(1-\alpha)^{2/(1+\alpha)}\,\eta_0^{-1/(1+\alpha)}L(t)^{1/(1+\alpha)} \end{aligned} \tag{21.21}$$

and

$$\begin{aligned} S(t+1) &= (1-u(t+1))\left(n^S(t+1)-1\right)L(t) \\ &= (1-u(t+1))\,\eta_1^{-1}L(t). \end{aligned} \tag{21.22}$$

Moreover, to satisfy (21.19) as equality, we need $(1-\alpha)^2\,U(t+1)^{-2\alpha}\,\eta_0^{-1} = X(t+1)\,\eta_1^{-1}$, or rearranging

$$X(t+1)\,\eta_1^{-1} = u(t+1)^{-2\alpha/(1+\alpha)}(1-\alpha)^{2(1-\alpha)/(1+\alpha)}\,\eta_0^{-(1-\alpha)/(1+\alpha)}L(t)^{-2\alpha/(1+\alpha)}. \tag{21.23}$$

Equilibrium dynamics are then determined by equations (21.21)-(21.23) together with (21.16). While the details of the behavior of this dynamical system are somewhat involved, the general picture is clear. Most interestingly, if an economy starts with both a low level of $X(0)$ and a low level of $L(0)$, but does not satisfy condition (21.20), then the economy will start in the Malthusian regime, only making use of the traditional technology and not investing in skills. As population increases wages fall, and at that point parents start finding it beneficial to invest in the skills of their children and firms start using the modern technology. Parents that invest in the skills of their children will typically have fewer children than parents rearing unskilled offspring (because η_1 is sufficiently larger than η_0, (21.17) will be greater than (21.18)). The aggregate rate of population growth and fertility are still high at first, but as the modern technology improves and the demand for skills increases, a larger fraction of parents start investing in the skills of their children and the rate of population growth declines. Ultimately, the rate of population growth approaches η_1^{-1}. This model thus gives a stylized representation of the demographic transition based on quality-quantity trade-off.

There are substantially richer models of the demographic transition in the literature. For example, there are many ways of introducing quality-quantity tradeoffs, and what spurs a change in the quality-quantity tradeoff may be an increase in capital intensity of production, changes in the wages of workers, or changes in the wages of women differentially affecting the desirability of market and home activities. Nevertheless, the general qualitative features are similar in that the quality-quantity tradeoff is often viewed as the major reason for the demographic transition. Despite this emphasis on the quality-quantity tradeoff, there is relatively little direct evidence that this tradeoff is important in general or in leading to the demographic transition. Other social scientists have suggested social norms, the large declines in mortality starting in the 19th century, or the reduced need for child labor as potential factors contributing to the demographic transition. As of yet, there is no general consensus on the causes of the demographic transition or on the role of the quality-quantity tradeoff in determining population dynamics. The study of population growth and demographic transition is an exciting and important area, and theoretical and empirical analyses of the factors affecting fertility decisions and how they interact with the reallocation of workers across different tasks (sectors) remain important and interesting questions to be explored.

21.3. Migration, Urbanization and The Dual Economy

Another major structural transformation over the process of development relates to changes in social and living arrangements. For example, as an economy develops, more individuals move from rural areas to cities and also undergo the social changes associated with separation from a small community and becoming part of a larger, more anonymous environment. Other social changes might also be important. For instance, certain social scientists regard the replacement of "collective responsibility systems" by "individual responsibility systems" as an important social transformation. This is clearly related to changes in the living arrangements of individuals (e.g., villages versus cities, or extended versus nuclear families). It is also linked to whether different types of contracts are being enforced by social norms and community enforcement, or whether they are enforced by legal institutions. There may also be a similar shift in the importance of the market, as more activities are mediated via prices rather than taking place inside the home or using the resources of an extended family or a broader community. This process of social change is both complex and interesting to study, though a detailed discussion of the literature and possible approaches to these issues falls beyond the scope of the current book.

Nevertheless, a brief discussion of some of these social changes are useful to illustrate other, more diverse facets of structural transformations associated with economic development. I illustrate the main ideas by focusing on the process of migration from rural areas and urbanization. Another reason to study migration and urbanization is that the reallocation of labor from rural to urban areas is closely related to the popular concept of *the dual economy*, which is an important theme of some of the older literature on development economics. According to this notion, less-developed economies consist of a modern sector and a traditional sector, but the connection between these two sectors is imperfect. The model of industrialization in the previous chapter (Section 20.3) featured a traditional and a modern sector, but these sectors traded their outputs and competed for labor in competitive markets. Dual economy approaches, instead, emphasize situations in which the traditional and the modern sectors function in parallel, but with only limited interactions. Moreover, the traditional sector is often viewed as less efficient than the modern sector, thus the lack of interaction may also be a way of shielding the traditional economy from its more efficient competitor. A natural implication of this approach will then be to view the process of development as one in which the less efficient traditional sector is replaced by the more efficient modern sector. Lack of development may in turn correspond to an inability to generate such reallocation.

In this section I first present a model of migration that builds on the work by Arthur Lewis (1954). A less-developed economy is modeled as a dual economy, with the traditional sector associated with villages and the modern sector with the cities. I then present a model inspired by Banerjee and Newman (1998) and Acemoglu and Zilibotti (1999), in which the traditional sector and the rural economy have a comparative advantage in community enforcement, even though in line with the other dual economy approaches, the modern economy (the city) enables the use of more efficient technologies. This model also illustrates how certain aspects of the traditional sector can shield the less productive firms from more productive competitors and slow down the process of development. Finally, I show how the import of technologies from more developed economies, along the lines of the models discussed in Section 18.4 of Chapter 18, may also lead to dual economy features as a byproduct of the introduction of more skill-intensive, modern technologies into less-developed economies.

21.3.1. Surplus Labor and the Dual Economy. Lewis argued that less-developed economies typically had *surplus labor*, that is, unemployed or underemployed labor, often in the villages. The dual economy can then be viewed as the juxtaposition of the modern sector, where workers are productively employed, with the traditional sector, where they are underemployed. The general tendency of less-developed economies to have lower levels of employment to population ratios was one of the motivations for Lewis's model. A key feature of Lewis's model is the presence of some barriers preventing, or slowing down, the allocation of workers away from the traditional sector towards urban areas, and the modern sector. I now present a reduced-form model that formalizes these notions.

Consider a continuous-time infinite-horizon economy that consists of two sectors or regions, which I refer to as urban and rural. Total population is normalized to 1. At time

$t = 0$, $L^U(0)$ individuals are in the urban area and $L^R(0) = 1 - L^U(0)$ are in the rural area. In the rural area, the only economic activity is agriculture and, for simplicity, suppose that the production function for agriculture is linear, thus total agricultural output is

$$Y^A(t) = B^A L^R(t),$$

where $B^A > 0$. In the urban area, the main economic activity is manufacturing. Manufacturing can only employ workers in the urban area and will employ all of the available workers. The production function therefore takes the form

$$Y^M(t) = F\left(K(t), L^U(t)\right),$$

where $K(t)$ is the capital stock, with initial condition $K(0)$. F is a standard neoclassical production function satisfying Assumptions 1 and 2. Let us also assume, for simplicity, that the manufacturing and the agricultural goods are perfect substitutes. Labor markets both in the rural and urban area are competitive. There is no technological change in either sector.

The key assumption is that because of barriers to mobility, there will only be slow migration of workers from rural to urban areas even when manufacturing wages are greater than rural wages. In particular, let us capture the dynamics in this model in a reduced-form way whereby capital accumulates only out of the savings of individuals in the urban area, thus

$$\dot{K}(t) = sF\left(K(t), L^U(t)\right) - \delta K(t), \tag{21.24}$$

where s is the exogenous saving rate and δ is the depreciation rate of capital. The important feature implied by this specification is that greater output in the modern sector leads to further accumulation of capital for the modern sector. An alternative, adopted in Section 20.3 of the previous chapter that will also be used in the next subsection, is to allow the size of the modern sector to directly influence its productivity growth, for example because of learning-by-doing externalities as in Romer (1986a) or because of endogenous technological change depending on the market size commanded by this sector (e.g., Exercise 20.19). For the purposes of the model in this subsection, which of these alternatives is adopted has no major consequences.

Given competitive labor markets, the wage rates in the urban and rural areas are

$$w^U(t) = \frac{\partial F\left(K(t), L^U(t)\right)}{\partial L} \text{ and } w^R(t) = B^A.$$

Let us assume that

$$\frac{\partial F(K(0), 1)}{\partial L} > B^A, \tag{21.25}$$

so that even if all workers are employed in the manufacturing sector at the initial capital stock, they will have a higher marginal product than working in agriculture.

Migration dynamics are assumed to take the following simple form:

$$\dot{L}^R(t) \begin{cases} = -\mu L^R(t) & \text{if } w^U(t) > w^R(t) \\ \in \left[-\mu L^R(t), 0\right] & \text{if } w^U(t) = w^R(t) \\ = 0 & \text{if } w^U(t) < w^R(t) \end{cases} \tag{21.26}$$

This equation implies that as long as wages in the urban sector are greater than those in the rural sector, there is a constant rate of migration. The speed of migration does not depend on the wage gap, which is an assumption adopted only to simplify the exposition. We may want to think of μ as small, so that there are barriers to migration and even when there are substantial gains to migrating to the cities, migration will take place slowly. When there is no wage gain to migrating, there will be no migration.

Now (21.25) implies that at date $t = 0$, there will be migration from the rural areas towards the cities. Moreover, assuming that $K(0)/L^U(0)$ is below the steady-state capital-labor ratio, the wage will remain high and will continue to attract further workers. To analyze this process in slightly greater detail, let us define

$$k(0) \equiv \frac{K(0)}{L^U(0)}$$

as the capital-labor ratio in manufacturing (the modern sector). As usual, let us also define the per capita production function in manufacturing as $f(k(t))$. Clearly, $w^U(t) = f(k(t)) - k(t) f'(k(t))$. Combining (21.24) and (21.26), we obtain that, as long as $f(k(t)) - k(t) f'(k(t)) > B^A$, the dynamics of this capital-labor ratio will be given by

$$\dot{k}(t) = sf(k(t)) - (\delta + \mu\nu(t)) k(t), \tag{21.27}$$

where $\nu(t) \equiv L^R(t)/L^U(t)$ is the ratio of the rural to urban population. Notice that when urban wages are greater than rural wages, the rate of migration, μ, times the ratio $\nu(t)$, plays the role of the rate of population growth in the basic Solow model. In contrast, when $f(k(t)) - k(t) f'(k(t)) \leq B^A$, there is no migration and

$$\dot{k}(t) = sf(k(t)) - \delta k(t). \tag{21.28}$$

Let us focus on the former case. Define the level of capital-labor ratio $\bar{k}$ such that urban and rural wages are equalized. This is given by

$$f(\bar{k}) - \bar{k} f'(\bar{k}) = B^A. \tag{21.29}$$

Once this level is reached, migration stops and $\nu(t)$ remains constant. After this level, equilibrium dynamics are given by (21.28). Therefore, the steady state must involve

$$\frac{sf(\hat{k})}{\hat{k}} = \delta. \tag{21.30}$$

For the analysis of transitional dynamics, which are our primary interest here, there are several cases to study. Let us focus on the one that appears most relevant for the experiences of many less-developed economies (leaving the rest to Exercise 21.4). In particular, suppose that the following conditions hold:

1. $k(0) < \hat{k}$, so that the economy starts with lower capital-labor ratio (in the urban sector) than the steady-state level. This assumption also implies that $sf(k(0)) - \delta k(0) > 0$.

2. $k(0) > \bar{k}$, which implies that $f(k(0)) - k(0) f'(k(0)) > B^A$, that is, wages are initially higher in the urban sector than in the rural sector.

3. $sf(k(0)) - (\delta + \mu\nu(0))\, k(0) < 0$, so that given the distribution of population between urban and rural areas, the initial migration will lead to a decline in the capital-labor ratio.

In this case, the economy starts with rural to urban migration at date $t = 0$. Since initially $\nu(0)$ is high, this migration reduces the capital-labor ratio in the urban area (which evolves according to the differential equation (21.27)). There are then two possibilities. In the first, the capital-labor ratio never falls below $\bar{k}$, thus rural to urban migration takes place at the maximum possible rate, μ, forever. Nevertheless, the effect of this migration on the urban capital-labor ratio is reduced over time as $\nu(t)$ declines with migration. Since we know that $sf(k(0)) - \delta k(0) > 0$, at some point the urban capital-labor ratio will start increasing, and it will eventually converge to the unique steady-state level $\hat{k}$. This convergence can take a long time, however, and notably, it is not necessarily monotone; the capital-labor ratio, and urban wages, first fall and then increase. The second possibility is that the initial surge in rural to urban migration reduces the capital-labor ratio to $\bar{k}$ at some point, say at date t'. When this happens, wages remain constant at B^A in both sectors and the rate of migration $\dot{L}^R(t)/L^R(t)$ adjusts exactly so that the capital-labor ratio remains at $\bar{k}$ for a while (recall that when urban and rural wages are equal, (21.26) admits any level of migration between zero and the maximum rate μ). In fact, the urban capital-labor ratio can remain at this level for an extended period of time. During this extended period of time, wages in both sectors remain stagnant. Ultimately, however, $\nu(t)$ will again decline sufficiently that the capital-labor ratio in the urban sector must start increasing. Once this happens, urban wages start increasing, migration takes place at the maximal rate μ and the economy again slowly converges to the capital-labor ratio $\hat{k}$ in the urban sector.

Therefore, this discussion illustrates how a simple model of migration can generate rich dynamics of population in rural and urban areas and also dynamics of wage difference between the modern and the traditional sectors.

The dynamics discussed above, especially in the first case, give the flavor of a *dual economy*. Wages and the marginal product of labor are higher in the urban area than in the rural area. If, in addition, μ is low, the allocation of workers from the rural to the urban areas will be slow, despite the higher wages. Thus the pattern of dual economy may be pronounced and may persist for a long time. It is also notable that rural to urban migration increases total output in the economy, because it enables workers to be allocated to activities in which their marginal product is higher. This process of migration increasing the output level in the economy also happens slowly because of the relatively slow process of migration.

The above discussion implies that, for the parameter configurations on which we have focused, the dual economy structure not only affects the social outlook of the society, which remains rural and agricultural for an extended period of time (especially when μ is small), but also leads to lower output than the economy could have generated by allocating labor more rapidly to the manufacturing sector. One should be cautious in referring to this as a "market failure," however, since we did not specify the reason why migration is slow.

21.3.2. Community Enforcement, Migration and Development. I now present a model inspired by Banerjee and Newman (1998) and Acemoglu and Zilibotti (1999). Banerjee and Newman consider an economy where the traditional sector has low productivity but is less affected by informational asymmetries and thus individuals can engage in borrowing and lending with limited monitoring and incentive costs. In contrast, the modern sector is more productive, but informational asymmetries create more severe credit market problems. Banerjee and Newman discuss how the process of development is associated with the reallocation of economic activity from the traditional to the modern sector and how this reallocation is slowed down by the informational advantage of the traditional sector. Acemoglu and Zilibotti (1999) view the development process as one of "information accumulation," and greater information enables individuals to write more sophisticated contracts and enter into more complex production relations. This process is then associated with changes in technology, changes in financial relations and social transformations, since greater availability of information and better contracts enable individuals to abandon less efficient and less information-dependent social and productive relationships.

The model in this subsection is simpler than both of these papers, but features a similar economic mechanism. Individuals who live in rural areas are subject to community enforcement. This means that they can enter into economic and social relationships without being unduly affected by moral hazard problems. When individuals move to cities, they can take part in more productive activities, but other enforcement systems are necessary to ensure compliance to social rules, contracts and norms. These systems will typically be associated with certain costs. As in the model of industrialization in the previous section, I also assume that the modern sector is subject to learning-by-doing externalities. Thus the productivity advantage of the modern sector grows as more individuals migrate to cities and work there. However, the community enforcement advantage of villages slows down this process.

Both labor markets are competitive and total population is normalized to 1. There are three differences from the model in the previous subsection. First, migration between the rural and urban areas is costless. Thus at any point in time an individual can switch from one sector to another. Second, instead of capital accumulation, there is an externality, so that output in the modern sector is given by

$$Y^M(t) = X(t)\, F\left(L^U(t), Z\right),$$

where $X(t)$ denotes the productivity of the modern sector, which will be determined endogenously via learning-by-doing externalities. In addition, Z denotes another factor of production in fixed supply (so that there are diminishing returns to labor), and the production function F satisfies Assumptions 1 and 2. The returns to factor Z are distributed back to individuals (and how they are distributed has no effect on the results). Moreover, let us assume that the technology in the modern sector evolves according to the differential equation

$$\dot{X}(t) = \eta L^U(t)\, X(t)^{\zeta},$$

where $\zeta \in (0,1)$. This equation builds in learning-by-doing externalities along the lines of Romer's (1986a) paper. The fact that $\zeta < 1$ implies that these externalities are less than those necessary for sustained growth.

Finally, let us also assume that rural areas have a comparative advantage in *community enforcement*. In particular, individuals engage in many social and economic activities, ranging from financial relations and employment to marriage and social relations. Many of these relationships in cities are anonymous and enforcement is through some type of monitoring by the law and relies on complex institutions. Such institutions often work imperfectly in most societies and particularly in less-developed economies. In contrast, rural areas house a small number of individuals who are typically engaged in long-term relationships. These long-term relationships enable the use of community enforcement in many activities. Thus with long-term relationships, individuals can pledge their reputation to borrow money, to obtain information about which individual would be most appropriate for a particular job or to ensure cooperation in other work or social relations. I represent these in a reduced-form way by assuming that, when in the urban area, an individual pays a flow cost of $\xi > 0$ due to imperfect monitoring and lack of community enforcement.

All individuals maximize the net present discounted value of their lifetime incomes. Since moving between urban and rural areas is costless, this implies that each individual should work in the sector that has the higher net wage at that time. This implies that in an interior equilibrium (where both the rural and the urban sectors are active), we must have

$$w^M(t) - \xi = w^A(t).$$

Competitive labor markets then imply that

$$w^M(t) = X(t)\frac{\partial F\left(L^U(t), Z\right)}{\partial L} \equiv X(t)\,\tilde{\phi}\left(L^U(t)\right),$$

where the second line defines the function $\tilde{\phi}$, which is strictly decreasing in view of Assumption 1 on the production function F. Substituting from the above relationships, labor market clearing implies that $X(t)\,\tilde{\phi}\left(L^U(t)\right) = B^A + \xi$, or

$$L^U(t) = \tilde{\phi}^{-1}\left(\frac{B^A+\xi}{X(t)}\right) \equiv \phi\left(\frac{X(t)}{B^A+\xi}\right),$$

where the second equality defines the function ϕ, which is strictly increasing in view of the fact that $\tilde{\phi}$ (and thus $\tilde{\phi}^{-1}$) is strictly decreasing. Therefore, the evolution of this economy can be represented by the differential equation

$$\dot{X}(t) = \eta\phi\left(\frac{X(t)}{B^A+\xi}\right)X(t)^{\zeta}.$$

A number of features about this law of motion are worth noting. First, the typical evolution of $X(t)$ will be given as in Figure 21.3, with an S-shaped pattern. This is because starting with a low initial value of $X(0)$, equilibrium in urban employment, $\phi\left(X(t)/\left(B^A+\xi\right)\right)$, will also be low during the early stages of development. This implies that there will be limited learning-by-doing and the modern sector technology will progress only slowly. However, as

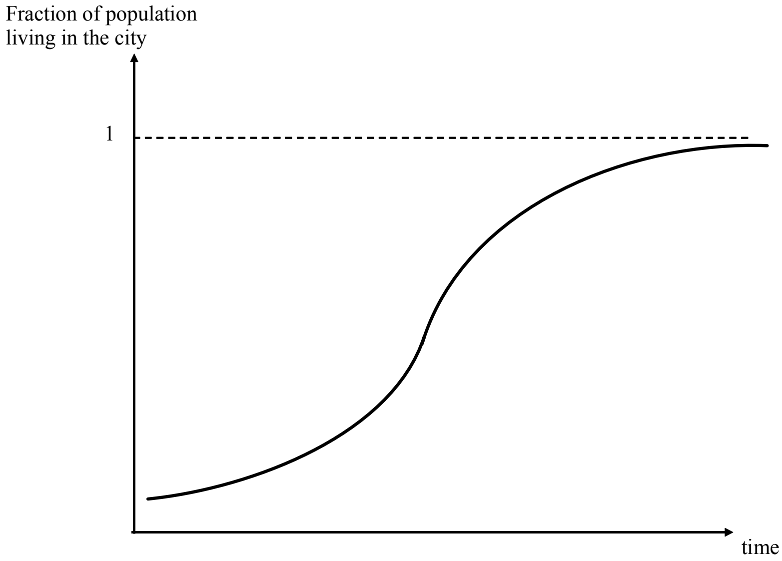

Figure 21.3. The dynamic behavior of the population in rural and urban areas.

$X(t)$ increases, $\phi\left(X(t)/\left(B^A+\xi\right)\right)$ also increases, raising the rate of technological change in the modern sector. Ultimately, however, $L^U(t)$ cannot exceed 1, so $\phi\left(X(t)/\left(B^A+\xi\right)\right)$ tends to a constant, and thus the rate of growth of X declines. Therefore, this reduced-form model generates an S-shaped pattern of technological change in the modern sector and an associated pattern of migration of workers from rural to urban areas.

Second and more importantly, the process of technological change in the modern sector and migration to the cities is slowed down by the comparative advantage of the rural areas in community enforcement. In particular, the greater is ξ, the slower is technological change and migration into urban areas. Since employment in the urban areas creates positive externalities, the community enforcement system in rural areas slows down the process of economic development in the economy as a whole. We may therefore conjecture that high levels of ξ, corresponding to greater community enforcement advantage of the traditional sector, will generally reduce growth and welfare in the economy. Counteracting this, however, are the static gains created by the better community enforcement system in rural areas. A high level of ξ will increase the initial level of consumption in the economy. Consequently, there is a tradeoff between dynamic and static welfare implications of different levels of ξ and this tradeoff is investigated formally in Exercise 21.5.

It is worth noting that differently from the model in subsection 21.3.1, there are no barriers to migration here, thus workers in the villages and cities receive the same wage. However, the functioning of the economy and the structure of social relations are different in these two areas. While villages and the rural economy rely on community enforcement, the city uses the modern technology and impersonal institutional checks in order to enforce various economic and social arrangements. Consequently, the dual economy in this model exhibits itself as much in the social dimension as in the economic dimension.

21.3.3. Inappropriate Technologies and the Dual Economy. I now discuss how ideas related to appropriate and inappropriate technologies presented in Chapter 18 may provide promising clues about other important aspects of the dual economy. Recall from Section 18.4 that less-developed economies often import their technologies from more advanced economies and that these technologies are typically designed for different factor proportions than those of the less-developed economy. For example, in Section 18.4, I emphasized the implications of a potential mismatch between the skills of the workforce of a less-developed economy and the skill requirements of modern technologies. However, in that model the equilibrium always involved all workers in the less-developed economy using the modern technology.

Here suppose that each technology is of the Leontief type, so that it requires a certain number of skilled and unskilled workers. For example, technology A_h will produce a total of $A_h L$ units of the unique final good, where L is the number of unskilled workers, but this technology requires a ratio of skilled to unskilled workers exactly equal to h (for example, the skilled workers will be the managers of the unskilled workers). Suppose A_h is increasing in h, so that more advanced technologies are more productive.

Now consider a less-developed economy that has access to all technologies A_h for $h \in [0, \bar{h}]$ for some $\bar{h} < \infty$. Suppose that the population of this economy consists of H skilled and L unskilled workers, such that $H/L < \bar{h}$. This inequality implies that not all workers can be employed with the most skill-intensive technology. What will the form of equilibrium be in this economy?

To answer this question, imagine that all markets are competitive, so that the allocation of workers to tasks will simply maximize output (recall the Second Welfare Theorem, Theorem 5.7). Then, the problem can be written as

$$\max_{[L(h)]_{h\in[0,\bar{h}]}} \int_0^{\bar{h}} A_h L(h)\, dh \tag{21.31}$$

subject to

$$\int_0^{\bar{h}} L(h)\, dh = L \text{ and } hL(h)\, dh = H,$$

where $L(h)$ is the number of unskilled workers assigned to work with technology A_h. The first-order conditions for this maximization problem can be written as

$$A_h \leq \lambda_L + h\lambda_H \text{ for all } h \in [0, \bar{h}], \tag{21.32}$$

where λ_L is the multiplier associated with the first constraint and λ_H is the multiplier associated with the second constraint. The first-order condition is written as an inequality, since not all technologies $h \in [0, \bar{h}]$ will be used, and those that are not active might satisfy this condition with a strict inequality.

Inspection of the first-order conditions implies that if $A_{\bar{h}}$ is sufficiently high and if $A_0 > 0$, the solution to this problem will have a simple feature. All skilled workers will be employed at technology $\bar{h}$, and together with them, there will be $L\left(\bar{h}\right) = H/\bar{h}$ unskilled workers employed with this technology. The remaining $L - L\left(\bar{h}\right)$ workers will be employed with the technology $h = 0$ (see Exercise 21.6). This equilibrium will then have the feature of a dual economy. Two very different technologies will be used for production, one more advanced (modern), and the other corresponding to the least advanced technology that is feasible. This dual economy structure emerges because of a non-convexity—to maximize output, it is necessary to operate the most advanced technology, but this exhausts all of the available skilled workers, implying that unskilled workers must be employed in technologies that do not require skilled inputs. This perspective therefore suggests that a dual economy structure may result from the import of technologies that are potentially mismatched with the supply of skills in an economy.

Models of dual economy based on this type of appropriate technology ideas have not been investigated in detail, though the literature on appropriate technology, which was discussed in Chapter 18, suggests that they may be important in practice. While this model focuses on the dual economy aspect in production, one can easily generalize this framework by assuming that the more advanced technology will be operated in urban areas and with contractual arrangements enforced by modern institutions, while the less advanced technology is operated in villages or rural areas. Thus models based on appropriate (or inappropriate) technology may be able to account for the broad patterns related to the dual economy, including rural to urban migration and changes in social arrangements.

21.4. Distance to the Frontier and Changes in the Organization of Production

In this section, I discuss how the structure of production changes over the process of development, and how this might be related both to changes in certain aspects of the internal organization of the firm and to a shift in the "growth strategy" of an economy—here, meaning whether the engine of growth is innovation or imitation. I illustrate these ideas using a simple model based on Acemoglu, Aghion and Zilibotti (2006). Because of space restrictions, I only provide a sketch of the model, mainly focusing on the production side.

Consider an economy that is behind the world technology frontier. There is no need to use country indices, since I focus on a single country, taking the behavior of the world technology frontier as given. Time is discrete and the economy is populated by two-period lived overlapping generations. Total population is normalized to 1. There is a unique final good, which is also taken as the numeraire. It is produced competitively using a continuum of machines with a technology similar to those in endogenous growth models in Chapter 14,

$$Y(t) = \int_0^1 A(\nu,t)^{\beta} x(\nu,t)^{1-\beta} \, d\nu, \tag{21.33}$$

where $A(\nu,t)$ is the productivity of machine variety ν at time t, $x(\nu,t)$ is the amount of this machine used in the production of the final good at time t, and $\beta \in (0,1)$.

Each machine good is produced by a monopolist $\nu \in [0,1]$ at a unit marginal cost in terms of the unique final good. The monopolist faces a competitive fringe of imitators that can copy its technology and also produce an identical machine with productivity $A(\nu,t)$, but can only do so at the cost of $\chi > 1$ units of final good. The existence of this competitive fringe forces the monopolist to charge a *limit price*:

$$p(\nu,t) = \chi > 1. \tag{21.34}$$

This limit price configuration will be an equilibrium when χ is not so high that the monopolist can set the unconstrained monopoly price. The condition for this is

$$\chi \leq 1/(1-\beta),$$

which I impose throughout. The parameter χ captures both technological factors and government regulations regarding competitive policy. A higher χ corresponds to a less competitive market. Given the demand implied by the final goods technology in (21.33) and the equilibrium limit price in (21.34), equilibrium monopoly profits are simply:

$$\pi(\nu,t) = \delta A(\nu,t), \tag{21.35}$$

where

$$\delta \equiv (\chi - 1)\, \chi^{-1/\beta} (1-\beta)^{1/\beta}$$

is a measure of the extent of monopoly power. In particular it can be verified that δ is increasing in χ for all $\chi \leq 1/(1-\beta)$.

In this model, the process of economic development will be driven not by capital accumulation—which was the force emphasized in some of the earlier models—but by technological progress, that is, by increases in $A(\nu,t)$. Let us assume that each monopolist $\nu \in [0,1]$ can increase its $A(\nu,t)$ by two complementary processes: (i) imitation (adoption of existing technologies); and (ii) innovation (discovery of new technologies). The key economic tradeoffs in the model arise from the fact that different economic arrangements (both in terms of the organization of firms and in terms of the growth strategy of the economy) will lead to different amounts of imitation and innovation.

To prepare for this point, let us define the average productivity of the economy in question at date t as:

$$A(t) \equiv \int_0^1 A(\nu,t)\, d\nu.$$

Let $\bar{A}(t)$ denote the productivity at the world technology frontier. The fact that this economy is behind the world technology frontier means that $A(t) \leq \bar{A}(t)$ for all t. The world technology frontier progresses according to the difference equation

$$\bar{A}(t) = (1+g)\, \bar{A}(t-1), \tag{21.36}$$

where the growth rate of the world technology frontier is taken to be

$$g \equiv \underline{\eta} + \bar{\gamma} - 1, \tag{21.37}$$

where $\underline{\eta}$ and $\bar{\gamma}$ will be defined below.

I assume that the process of imitation and innovation leads to the following law of motion of each monopolist's productivity:

$$A(\nu, t) = \eta \bar{A}(t-1) + \gamma A(t-1) + \varepsilon(\nu, t), \tag{21.38}$$

where $\eta > 0$ and $\gamma > 0$, and $\varepsilon(\nu, t)$ is a random variable with zero mean, capturing differences in innovation performance across firms and sectors.

In equation (21.38), $\eta \bar{A}(t-1)$ stands for advances in productivity coming from *adoption* of technologies from the frontier (and thus depends on the productivity level of the frontier, $\bar{A}(t-1)$), while $\gamma A(t-1)$ stands for the component of productivity growth coming from innovation (building on the existing knowledge stock of the economy in question at time $t-1$, $A(t-1)$). Let us also define

$$a(t) \equiv \frac{A(t)}{\bar{A}(t)}$$

as the (inverse) measure of the country's *distance to the technological frontier* at date t.

Now, integrate (21.38) over $\nu \in [0,1]$, use the fact that $\varepsilon(\nu, t)$ has mean zero, divide both sides by $\bar{A}(t)$ and use (21.36) to obtain a simple linear relationship between a country's distance to frontier $a(t)$ at date t and the distance to frontier $a(t-1)$ at date $t-1$:

$$a(t) = \frac{1}{1+g}(\eta + \gamma a(t-1)). \tag{21.39}$$

This equation is similar to the technological catch-up equation in Section 18.2 in Chapter 18. It shows how the dual process of imitation and innovation may lead to a process of convergence. In particular, as long as $\gamma < 1 + g$, equation (21.39) implies that $a(t)$ will eventually converge to 1. Second, the equation also shows that the relative importance of imitation and innovation depends on the distance to the frontier of the economy in question. In particular when $a(t)$ is large (meaning the country is close to the frontier), innovation, γ, matters more for growth. In contrast when $a(t)$ is small (meaning the country is farther from the frontier), imitation, η, is relatively more important.

To obtain further insights, let us now endogenize η and γ using a reduced-form approach. Following the analysis in Acemoglu, Aghion and Zilibotti (2006), I model the parameters η and γ as functions of the investments undertaken by the entrepreneurs and the contractual arrangement between firms and entrepreneurs. The key idea is that there are two types of entrepreneurs: high-skill and low-skill. When an entrepreneur starts a business, his skill level is unknown and is revealed over time through his subsequent performance. This implies that there are two types of "growth strategies" that are possible. The first one emphasizes *selection* of high-skill entrepreneurs and will replace any entrepreneur that is revealed to be low skill. This growth strategy will involve a high degree of churning (creative destruction) and a large number of young entrepreneurs (as older unsuccessful entrepreneurs are replaced by new young entrepreneurs). The second strategy maintains experienced entrepreneurs in place even when they have low skills. This strategy therefore involves an organization of firms relying on "longer-term relationships" (here between entrepreneurs and the credit market), an emphasis on experience and cumulative earnings, and less creative destruction. While low-skill entrepreneurs are less productive than high-skill entrepreneurs, there are potential reasons for why an experienced low-skill entrepreneur might be preferred to a new young entrepreneur. For example, experience may increase productivity, at least in certain tasks. Alternatively, Acemoglu, Aghion and Zilibotti (2006) show that in the presence of credit market imperfections, the retained earnings of an old entrepreneur may provide him with an advantage in the credit market (because he can leverage his existing earnings to raise more money and undertake greater productivity-enhancing investments). I denote the strategy based on selection by $R = 0$, while the strategy that maintains experienced entrepreneurs in place is denoted by $R = 1$.

The key reduced-form assumption here will be that experienced entrepreneurs (either because of the value of experience or because of their retained earnings) are better at increasing the productivity of their company when this involves the imitation of technologies from the world frontier, which can be thought to correspond to relatively "routine" tasks. High-skill entrepreneurs, on the other hand, are more innovative and generate higher growth due to innovation. Thus the tradeoff between $R = 1$ and $R = 0$ and the associated tradeoff between organizational forms boils down to the tradeoff between imitation of technologies from the world technology frontier versus innovation. For this reason, I refer to the first one as *imitation-based growth strategy* and to the second one as *innovation-based growth strategy.* Motivated by these considerations, let us assume that the equation for the law of motion of the distance to frontier, (21.39), takes the form

$$a(t) = \begin{cases} \frac{1}{1+g}(\bar{\eta} + \underline{\gamma} a(t-1)) & \text{if} \quad R(t) = 1 \\ \\ \frac{1}{1+g}(\underline{\eta} + \bar{\gamma} a(t-1)) & \text{if} \quad R(t) = 0. \end{cases} \tag{21.40}$$

Let us also impose

$$\underline{\gamma} < \bar{\gamma} < 1 + g \text{ and } \bar{\eta} > \underline{\eta}. \tag{21.41}$$

The first part of this assumption follows immediately from the notion that high-skill entrepreneurs are better at innovation, while the second part, in particular, that $\bar{\gamma} > \underline{\gamma}$, builds in the feature that experienced entrepreneurs are better at imitation. When the imitation-based growth strategy is pursued, experienced entrepreneurs are not replaced, and consequently, there is greater transfer of technology from the world technology frontier. The final part of this assumption, $\underline{\gamma} < 1 + g$, simply ensures that imitation-based growth will not lead to faster growth than the world technology frontier. We can thus interpret Assumption (21.37) as stating that the world technology frontier advances due to innovation-based growth strategy, which is natural since a country at the world technology frontier cannot imitate from others.

Figure 21.4 draws equation (21.40), and shows that the economy with long-term contracts ($R = 1$) achieves greater growth (higher level of $a(t)$ for given $a(t-1)$) through the imitation channel, but lower growth through the innovation channel. The figure also shows that which regime maximizes the growth rate of the economy depends on the level of $a(t-1)$, that is, on the distance of the economy to the world technology frontier. In particular, inspection of

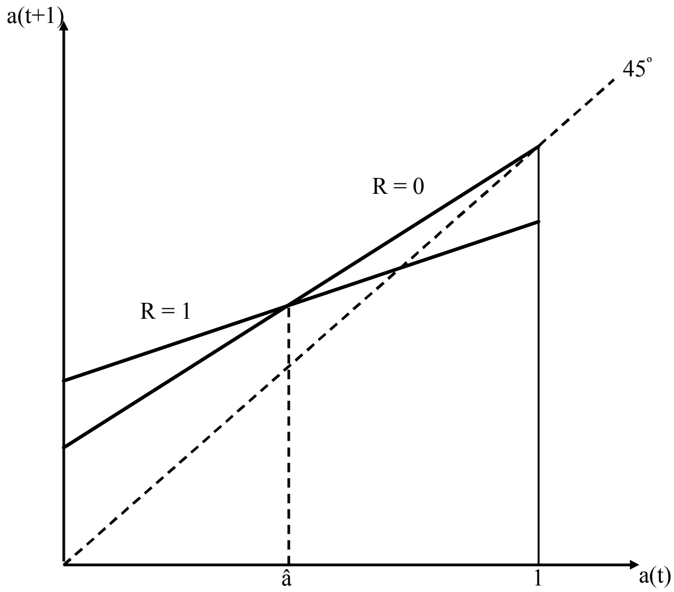

Figure 21.4. The growth-maximizing threshold and the dynamics of the distance to frontier in the growth-maximizing equilibrium.

(21.40) is sufficient to establish that there exists a threshold

$$\hat{a} \equiv \frac{\bar{\eta} - \eta}{\bar{\gamma} - \underline{\gamma}} \in (0, 1) \tag{21.42}$$

such that when $a(t-1) < \hat{a}$, the imitation-based strategy, $R = 1$ leads to greater growth, and when $a(t-1) > \hat{a}$, the innovation-based strategy, $R = 0$, achieves higher growth. Thus if the economy were to pursue a growth-maximizing sequence of strategies, it would start with $R = 1$ and then switch to an innovation-based strategy, $R = 0$, once it is sufficiently close to the world technology frontier. In the imitation-based regime, incumbent entrepreneurs are sheltered from the competition of younger entrepreneurs and this may enable the economy to make better use of the experience of older entrepreneurs or to finance greater investments out of their retained earnings. In contrast, the innovation-based regime is based on an organizational form relying on greater selection of entrepreneurs and places greater emphasis on maximizing innovation at the expense of experience, imitation and investment.

Figure 21.4 describes the law of motion of technology in an economy as a function of the organization of firms (markets), captured by R. It does not specify what the equilibrium sequence of $\{R(t)\}_{t=0}^{\infty}$ is. To determine this sequence, we need to specify the equilibrium behavior, which involves the selection of entrepreneurs as well as the functioning of credit markets. Space restrictions preclude me from providing a full analysis of the equilibrium in such a model. Instead, I informally discuss some of the main insights of such an analysis.

Conceptually, one might want to distinguish among four configurations, which arise as equilibria under different institutional settings and parameter values.

1. *Growth-maximizing equilibrium*: the first and the most obvious possibility is an equilibrium that is growth maximizing. In particular, if markets and entrepreneurs have growth maximization as their objective and are able to solve the agency problems, have the right decision-making horizon and are able to internalize the pecuniary and non-pecuniary externalities, we would obtain an efficient equilibrium. This equilibrium will take a simple form:

$$R(t) = \begin{cases} 1 & \text{if} \quad a(t-1) < \hat{a} \\ 0 & \text{if} \quad a(t-1) \geq \hat{a} \end{cases}$$

so that the economy achieves the upper envelope of the two lines in Figure 21.4. In this case, there is no possibility of outside intervention to increase the growth rate of the economy.[2] Moreover, an economy starting with $a(0) < 1$ always achieves a growth rate greater than g, and will ultimately converge to the world technology frontier, that is, $a(t) \to 1$. In this growth-maximizing equilibrium, the economy first starts with a particular set of organizations/institutions, corresponding to $R = 1$. Then, the economy undergoes a structural transformation—in this case, a change in its organizational form—switching from $R = 1$ to $R = 0$. In our simple economy, this structural transformation takes the form of long-term relationships disappearing and being replaced by shorter-term relationships, by greater competition among entrepreneurs and firms and by better selection of entrepreneurs.

2. *Underinvestment equilibrium*: the second potential equilibrium configuration involves the following equilibrium organizational form:

$$R(t) = \begin{cases} 1 & \text{if} \quad a(t-1) < a_r(\delta) \\ 0 & \text{if} \quad a(t-1) \geq a_r(\delta) \end{cases}$$

where $a_r(\delta) < \hat{a}$. Figure 21.5 depicts this visually, with the thick black lines corresponding to the equilibrium law of motion of the distance to the frontier, a. How is $a_r(\delta)$ determined? Acemoglu, Aghion and Zilibotti (2006) show that when investment is important for innovation and credit markets are imperfect, then the retained earnings of old (experienced) entrepreneurs enable them to undertake greater investments. However, because of monopolistic competition, there is the standard *appropriability effect*, whereby an entrepreneur that undertakes a greater investment does not capture all the surplus generated by this investment because some of it accrues to households in the form of greater consumer surplus. The appropriability effect discourages investments, and in this context because greater investments are associated with more experienced, older entrepreneurs, it discourages the imitation-based strategy. This description also explains why this equilibrium is referred to as the "underinvestment equilibrium"; in the range $a \in (a_r(\delta), \hat{a})$, the economy could reach a higher growth rate (as shown in the figure) by choosing $R(t) = 1$, but because the appropriability effect

[2] However, recall that growth-maximization is not necessarily the same as welfare-maximization. Depending on how preferences and investments are specified, the growth-maximizing allocation may not be welfare-maximizing.

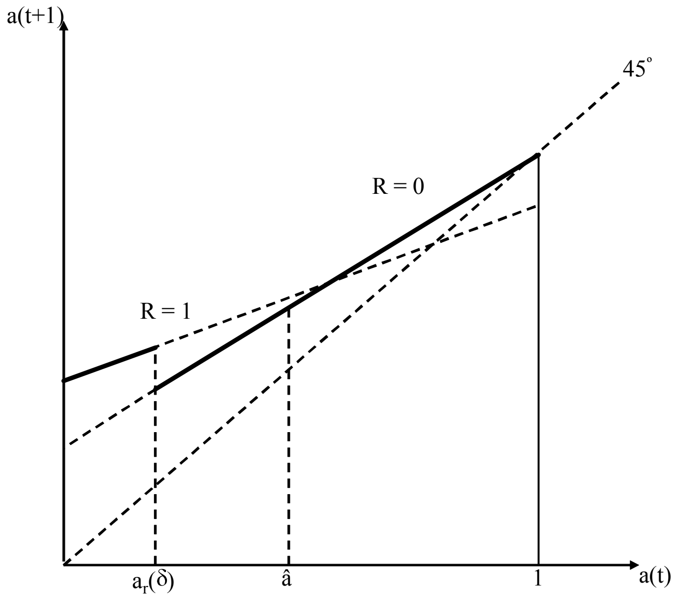

Figure 21.5. Dynamics of the distance to frontier in the underinvestment equilibrium.

discourages investments, there is a switch to the innovation-based equilibrium earlier than the growth-maximizing threshold.

A notable feature is that although the equilibrium is different from the previous case, it again starts with $R = 1$ and is followed by a structural transformation, that is, by a switch to the innovation-based regime ($R = 0$). Moreover, the economy still ultimately converges to the world technology frontier, that is, $a(t) = 1$ is reached as $t \to \infty$. The only difference is that the structural transformation from $R = 1$ to $R = 0$ happens too soon, at $a(t-1) = a_r(\delta)$, rather than at the growth-maximizing threshold $\hat{a}$.

Consequently, in this case, a *temporary* government intervention may increase the growth rate of the economy. The temporary aspect is important here, since the best that the government can do is to increase the growth rate while $a \in (a_r(\delta), \hat{a})$. How can the government achieve this? Subsidies to investment would be one possibility. Acemoglu, Aghion and Zilibotti (2006) show that the degree of competition in the product market also has an indirect effect on the equilibrium, as emphasized by the notation $a_r(\delta)$. In particular, a higher level of δ, which corresponds to lower competition in the product market (higher χ), will increase $a_r(\delta)$, and thus may close the gap between $a_r(\delta)$ and $\hat{a}$. Nevertheless, it has to be noted that reducing competition will create other, static distortions (because of higher markups). Moreover and more importantly, we will see in the next two configurations that reducing competition can have much more detrimental effects on economic growth, so any use of competition policy for this purpose must be subject to serious caveats.

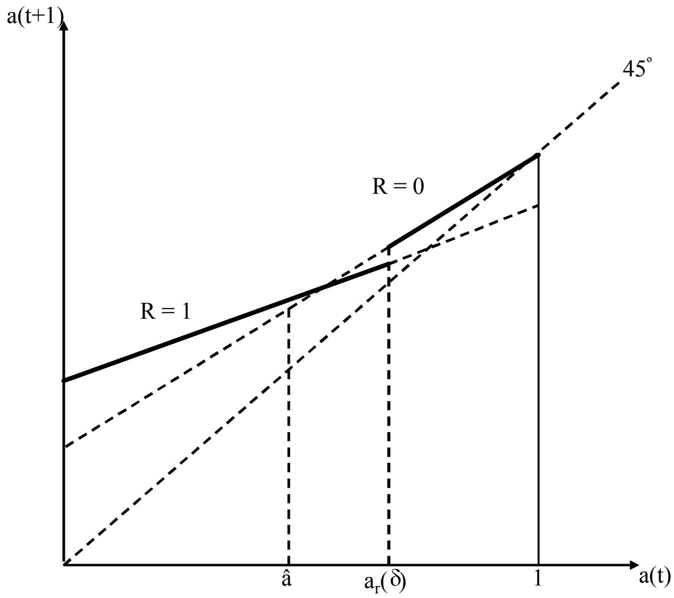

Figure 21.6. Dynamics of the distance to frontier in the sclerotic equilibrium.

3. *Sclerotic equilibrium*: the third possibility is a sclerotic equilibrium in which $a_r(\delta) > \hat{a}$, so that low-productivity incumbents survive even when they are potentially damaging to economic growth. Acemoglu, Aghion and Zilibotti (2006) show that this configuration can also arise in equilibrium because the retained earnings of incumbent entrepreneurs act as a *shield* protecting them against the creative destruction forces brought about by new entrepreneurs. Consequently, in general, the retained earnings or other advantages of experienced entrepreneurs both have (social) benefits and costs, and which of these will dominate depends on parameter values. When the benefits dominate, the equilibrium may feature too rapid a switch to the innovation-based strategy, and when the costs dominate, the economy may experience sclerosis in the imitation regime, with excessive protection of incumbents.

The resulting pattern in this case is drawn in Figure 21.6. Now the economy fails to achieve the maximum growth rate for a range of values of a such that $a \in (\hat{a}, a_r(\delta))$. In this range, the innovation-based regime would be growth-maximizing, but the economy is stuck with the imitation-based regime because the retained earnings and the power of the incumbents prevent the transition to the more efficient organizational forms. An interesting feature is that, as Figure 21.6 shows, this economy also follows a pattern in line with Kuznets's vision; it starts with a distinct set of organizations, represented by $R = 1$, and then switches to a different set of arrangements, $R = 0$. Like the previous two types of equilibria, this case also features convergence to the world technology frontier, that is, to $a = 1$.

4. *Non-convergence trap equilibrium:* the fourth possibility is related to the third one and also involves $a_r(\delta) > \hat{a}$. However, now the gap between $a_r(\delta)$ and $\hat{a}$ is larger as depicted

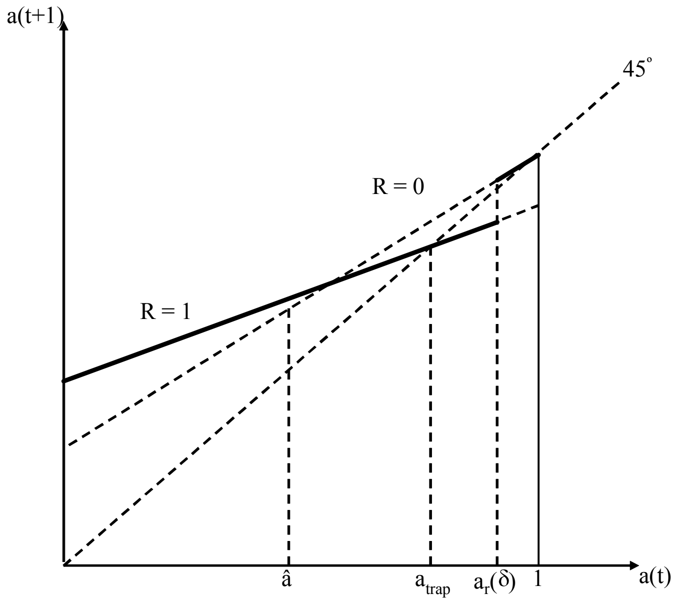

Figure 21.7. Dynamics of the distance to frontier in a non-convergence trap. If the economy starts with $a(0) < a_{trap}$, it fails to converge to the world technology frontier and instead converges to a_{trap}.

in Figure 21.7, and includes the level of a, a_{trap}, such that

$$a_{trap} \equiv \frac{\bar{\eta}}{1+g-\underline{\gamma}}.$$

Inspection of (21.40) immediately reveals that if $a(t-1) = a_{trap}$ and $R(t) = 1$, the economy will remain at a_{trap}. Therefore, in this case, the retained earnings or the experience of incumbent firms afford them so much protection that the economy never transitions to the innovation-based equilibrium. This not only retards growth for a temporary interval, but also pushes the economy into a non-convergence trap. In particular, this is the only equilibrium pattern in which the economy fails to converge to the frontier; with the imitation-based regime, $R = 1$, the economy does not grow beyond a_{trap}, and at this distance to frontier, the equilibrium keeps choosing $R = 1$.

This equilibrium therefore illustrates the most dangerous scenario—that of non-convergence. Encouraging imitation-based growth, for example by supporting incumbent firms, may at first appear as a good policy. But in practice, it may condemn the economy to non-convergence. This is also the only case in which the switch to $R = 0$ and the associated structural transformation dos *not* occur because the economy remains trapped. In many ways, this is in line with Kuznets' vision; the resulting economy is an underdeveloped one, partly because it is unable to realize the structural transformation necessary for the process of economic development.

Taken together the four scenarios suggest that depending on the details of the model, there should be no presumption that the efficient or the growth-maximizing sequence of growth strategies will be pursued. Thus, some degree of government intervention might be useful. However, the third and the fourth cases also emphasize that government intervention can have negative unintended consequences. It may improve growth during a limited period of time (in the second scenario this will be when $a \in (a_r(\delta), \hat{a})$), but it may subsequently create much more substantial costs by leading to a non-convergence trap as shown in Figure 21.7.

Even though the implications of these four scenarios for government intervention are mixed, their implications for changes in the structure of organizations over the development process are clearer; regardless of which scenario applies, the economy starts with a distinct organization of production, where longer-term contracts, incumbent producers, experience and imitation are more important, and then, except in the non-convergence trap equilibrium, it ultimately switches to an equilibrium with greater creative destruction, shorter-term relationships, younger entrepreneurs and more innovation. This is another facet of the structural transformations emphasized by Kuznets as part of the process of economic development. The framework presented here, though reduced-form, can also be used to study other aspects of changes in the organization of production over the process of development (see Exercise 21.7).

21.5. Multiple Equilibria From Aggregate Demand Externalities and the Big Push

I now present a simple model of multiple equilibria arising from aggregate demand externalities, based on Murphy, Shleifer and Vishny's (1989) "big push" model. This model formalizes ideas first proposed by Rosenstein-Rodan (1943), Hirschman (1958) and Nurske (1958), that economic development can (or should) be viewed as a move from one (Pareto inefficient) equilibrium to another, more efficient equilibrium. Moreover, these early development economists argued that this type of move requires coordination among different individuals and firms in the economy, thus a *big push*. As already discussed in Chapter 4, multiple equilibria, literally interpreted, are unlikely to be the root cause of persistently low levels of development, since if there is indeed a Pareto improvement—a change that will make *all* individuals better-off—it is unlikely that the necessary coordination cannot be achieved for decades or even centuries. Nevertheless, the forces leading to multiple equilibria highlight important economic mechanisms that can be associated with market failures slowing down, or even preventing, the process of development. Moreover, dynamic versions of models of multiple equilibria can lead to multiple state states, whereby once an economy ends up in a steady state with low economic activity, it may get stuck there (and there is no possibility of a coordination to jump to the other steady state). Models with multiple steady states, which are more useful for thinking about the process of long-run development than models with multiple equilibria, will be discussed in the next section.

Murphy, Shleifer and Vishny formalize the ideas related to the big push using a model with multiple equilibria due to aggregate demand externalities. The economy has two periods, $t = 1$ and 2, and admits a representative household with preferences

$$U(C(1), C(2)) = \frac{C(1)^{1-\theta} - 1}{1-\theta} + \beta \frac{C(2)^{1-\theta} - 1}{1-\theta}$$

where $C(1)$ and $C(2)$ denote consumption at the two dates; β is the discount factor of the households; and θ plays a similar role to before; $1/\theta$ is the intertemporal elasticity of substitution and determines how willing individuals are to substitute consumption between date 1 and date 2. The representative household supplies labor inelastically and the total labor supply is denoted by L.

The resource constraints are

$$C(1) + I(1) \leq Y(1) \text{ and } C(2) \leq Y(2), \tag{21.43}$$

where $I(1)$ denotes investment in the first date, $Y(t)$ is total output at date t, and investment is only possible in the first date.

Households can borrow and lend, so their budget constraint can be represented as

$$C(1) + \frac{C(2)}{1+r} \leq w(1) + \pi(1) + \frac{w(2) + \pi(2)}{1+r},$$

where $\pi(t)$ denotes the profits accruing to the representative household, and $w(t)$ is the wage rate at time t. r is the interest rate between periods 1 and 2 and will adjust in equilibrium so that the aggregate the resource constraints, (21.43), hold.

The final good is produced from a CES aggregate of differentiated intermediate goods, with production function

$$Y(t) = \left(\int_0^1 y(\nu, t)^{\frac{\varepsilon-1}{\varepsilon}} d\nu\right)^{\frac{\varepsilon}{\varepsilon-1}},$$

where $y(\nu, t)$ is the output level of intermediate ν at date t. As usual ε is the elasticity of substitution between intermediate goods and we assume that $\varepsilon > 1$.

The production functions of intermediate goods in the two periods are as follows:

$$y(\nu, 1) = l(\nu, 1)$$

$$y(\nu, 2) = \begin{cases} l(\nu, 2) & \text{with old technology} \\ \alpha l(\nu, 2) & \text{with new technology} \end{cases} \tag{21.44}$$

where $\alpha > 1$ and $l(\nu, t)$ denotes labor devoted to the production of intermediate good ν at time t. Labor market clearing requires

$$\int_0^1 l(\nu, t)\, d\nu \leq L. \tag{21.45}$$

At date 1, there is a designated producer for each intermediate, which I also refer to as a "monopolist". A competitive fringe of firms can also enter and produce each good as productively as the designated producer. At date 1, the designated producer can also invest in the new technology, which costs F in terms of the final good. If this investment is undertaken, this producer's productivity at date 2 will be higher by a factor $\alpha > 1$ as indicated by equation (21.44). In contrast, the fringe will not benefit from this technological improvement, thus the designated producer will have some degree of monopoly power. The profits from intermediate producers are naturally allocated to the representative household.

We will look for the pure-strategy symmetric Subgame Perfect Equilibria (SSPE) of this two-period economy (see, for example, Section 18.5 in Chapter 18). SSPE is defined in the usual fashion, as a combination of production and investment decisions for firms and consumption decisions for households that are best responses to each other in both periods.

First, since all goods are symmetric, the first period labor market clearing is straightforward and requires

$$l(\nu,1) = L \text{ for all } \nu \in [0,1]$$

(recall that the measure of sectors and firms is normalized to 1). This implies that

$$Y(1) = L.$$

At date 2, the equilibrium will depend on how many firms have adopted the new technology. Since the focus is on SSPE, it is sufficient to consider the two extreme allocations, where all firms adopt the new technology and where no firm adopts. In either case, the marginal productivity of all sectors is the same, so labor will be allocated equally so that

$$l(\nu,2) = L \text{ for all } \nu \in [0,1].$$

Consequently, when the technology is not adopted,

$$Y(2) = L$$

and when the technology is adopted by all the firms,

$$Y(2) = \alpha L.$$

I now turn to the pricing decisions. In the first date, the designated producers have no monopoly power because of the competitive fringe, thus they charge price equal to marginal cost, which is $w(1)$, and they make zero profits. Since total output is equal to $Y(1) = L$, this also implies that the equilibrium wage rate is $w(1) = 1$. In the second date, if the technology is not adopted, the equilibrium is identical to that at date 1, so $w(2) = 1$, and thus there will be no profits. In this case there is also no investment, so consumption at both dates is equal to L. Since the consumption Euler equation is $C(1)^{-\theta} = (1+r)\beta C(2)^{-\theta}$, the equilibrium interest rate in this case is

$$\hat{r} = \beta^{-1} - 1. \tag{21.46}$$

Next consider the situation in which the designated producers have invested in the advanced technology. Now they can produce α units of output with one unit of labor, while the fringe of competitive firms still produces one unit of output with one unit of labor. This implies that the designated producers have some monopoly power. The extent of this monopoly power depends on the comparison of ε and α.

Let us first determine the demand facing each producer, which is given as a solution to the following program of profit maximization for the final good sector:

$$\max_{[y(\nu,2)]_{\nu\in[0,1]}} \left[\int_0^1 y(\nu,2)^{\frac{\varepsilon-1}{\varepsilon}}\,d\nu\right]^{\frac{\varepsilon}{\varepsilon-1}} - \int_0^1 p(\nu,2)\,y(\nu,2)\,d\nu,$$

where $p(\nu,2)$ is the price of intermediate ν at date 2. The first-order condition to this program implies

$$y(\nu,2) = p(\nu,2)^{-\varepsilon}\,Y(2)\,. \tag{21.47}$$

This expression is useful in laying the foundations for the aggregate demand externalities; the demand for intermediate ν depends on the total amount of production, $Y(2)$. The familiar feature of the demand curve (21.47) is that it is isoelastic. To make further progress, first imagine the situation in which there is no fringe of competitive producers. In that case, each designated producer will act as an unconstrained monopolist and maximize its profits given by price minus marginal cost times quantity, that is,

$$\pi(\nu,2) = \left(p(\nu,2) - \frac{w(2)}{\alpha}\right) y(\nu,2)\,.$$

Substituting from (21.47), the firm maximization problem is thus

$$\max_{p(\nu,2)} \left(p(\nu,2) - \frac{w(2)}{\alpha}\right) p(\nu,2)^{-\varepsilon}\,Y(2)\,,$$

which gives the profit-maximizing price as

$$p(\nu,2) = \frac{\varepsilon}{\varepsilon-1}\frac{w(2)}{\alpha}.$$

This is the standard monopoly price formula with a markup related to demand elasticity over the marginal cost. The markup is constant because the demand elasticity is constant.

However, the monopolist can only charge this price if the competitive fringe cannot enter and make profits stealing the entire market at this price. Since the competitive fringe can produce one unit using one unit of labor, the monopolist can only charge this price if $\varepsilon/\left((\varepsilon-1)\,\alpha\right) \le 1$. Otherwise, the price would be too high and the competitive fringe would enter. Let us assume that α is not so high as to make the monopolist unconstrained. In other words, let us assume

$$\frac{\varepsilon}{\varepsilon-1}\frac{1}{\alpha} > 1. \tag{21.48}$$

Under this assumption, the monopolist will be forced to charge a *limit price*. It is straightforward to see that this equilibrium limit price would be $p^* = w(2)$. Consequently, given (21.48), each monopolist would make per unit profits equal to

$$w(2) - \frac{w(2)}{\alpha} = \frac{\alpha-1}{\alpha}w(2)\,.$$

Total profits are then obtained from (21.47) as

$$\pi(2) = \frac{\alpha-1}{\alpha}w(2)^{1-\varepsilon}\,Y(2)\,. \tag{21.49}$$

The wage rate can be determined from income accounting. Total production will be equal to $Y(2) = \alpha L$, and this has to be distributed between profits and wages, thus

$$\frac{\alpha - 1}{\alpha} w(2)^{1-\varepsilon} \alpha L + w(2) L = \alpha L,$$

which has a solution of $w(2) = 1$, exactly the same as in the case without the technological investments. Intuitively, wages in this economy are determined by the demand from the competitive fringe and thus the greater marginal product does not directly benefit workers. Instead, it increases monopolists' profits. Nevertheless, all of these profits are redistributed to households, who are the owners of the firms. Thus $C(2) = \alpha L$. However, because of investment in the new technology at date 1, $C(1) = L - F$. The consumption Euler equation now requires $(L - F)^{-\theta} = (1 + \tilde{r}) \beta (\alpha L)^{-\theta}$, which gives the equilibrium interest-rate in this case as

$$\tilde{r} = \beta^{-1} \left(\frac{\alpha L}{L - F} \right)^{\theta} - 1 > \hat{r},$$

where $\hat{r}$ is given by (21.46). The interest rate in this case, $\tilde{r}$, is higher than $\hat{r}$ because individuals are now being asked to forgo date 1 consumption for date 2 consumption. Note also that the greater is θ, the higher is $\tilde{r}$, since with a greater θ, there is less intertemporal substitution. Also a higher F, meaning a greater consumption sacrifice at date 1, implies a higher interest rate.

The question is whether a monopolist will find it profitable to undertake the investment at date 1. The reason for the possibility of multiplicity of equilibria is that the answer to this question will depend on whether other firms are undertaking the investment or not. Let us first consider a situation in which no other firm is undertaking the investment, and consider the incentives of a single monopolist to undertake such an investment.

In this case total output at date 2 is equal to L (since the firm considering investment is infinitesimal), and the market interest rate is given by $\hat{r}$. Moreover, from (21.49) and the fact that $w(2) = 1$, profits at date 2 are

$$\pi^N(2) = \frac{\alpha - 1}{\alpha} L,$$

where the superscript N denotes that no other firm is undertaking the investment. Therefore, the net discounted profits at date 1 for the firm in question are

$$\Delta \pi^N = -F + \frac{1}{1 + \hat{r}} \frac{\alpha - 1}{\alpha} L = -F + \beta \frac{\alpha - 1}{\alpha} L.$$

Next consider the case in which all other firms are undertaking the investment. In this case, profits at date 2 are

$$\pi^I(2) = (\alpha - 1) L,$$

where the superscript I designates that all other firms are undertaking the investment. Consequently, the profit gain from investing at date 1 is

$$\Delta \pi^I = -F + \frac{1}{1 + \tilde{r}} (\alpha - 1) L = -F + \beta \left(\frac{\alpha L}{L - F} \right)^{-\theta} (\alpha - 1) L.$$

As discussed above, the idea of the paper by Murphy, Shleifer and Vishny (1989), similar to the ideas of many economists writing on economic development before them, was to generate multiple equilibria, with one corresponding to backwardness and the other to industrialization. In this context, this means that for the same parameter values, both the allocations with no investment in the new technology and with all monopolists investing in the new technology should be equilibria. This is possible if

$$\Delta\pi^N < 0 \text{ and } \Delta\pi^I > 0, \tag{21.50}$$

that is, when nobody else invests, investment is not profitable, and when all other firms invest, investment is profitable. This is clearly possible because the *aggregate demand externalities* ensure that $\pi^I > \pi^N$; when other firms invest, they produce more, there is greater aggregate demand, and profits from the new technology are higher. Counteracting this effect is the fact that the interest rate is also higher when all firms invest. Therefore, the existence of multiple equilibria requires the interest rate effect not to be too strong. For example, in the extreme case where preferences are linear ($\theta = 0$), we have

$$\Delta\pi^I = -F + \beta(\alpha - 1)L > \Delta\pi^N = -F + \beta\frac{\alpha-1}{\alpha}L,$$

so the configuration in (21.50) is certainly possible. More generally, the condition for the existence of multiple equilibria is that:

$$\beta\left(\frac{\alpha L}{L-F}\right)^{-\theta}(\alpha - 1)L > F > \beta\frac{\alpha-1}{\alpha}L. \tag{21.51}$$

It is also straightforward to see that whenever both equilibria exist, the equilibrium with investment Pareto dominates the one without investment, since condition (21.51) implies that all households are better-off with the upward sloping consumption profile giving them higher consumption at date 2 (see Exercise 21.8). Therefore, this analysis establishes that when condition (21.51) is satisfied, there will exist two SSPE. In one of these, all firms undertake the investment at date 1 and households are better-off, while in the other one there are no investments in new technology and greater market failures. Intuitively, investing in the new technology at date 1 is profitable only when there is sufficient aggregate demand at date 2, and in turn, there will be sufficient demand at date 2 when all firms invest in the new technology. Aggregate demand externalities are responsible for multiple equilibria here. In particular, the investment decision of a firm creates a positive (pecuniary) externality on other firms by increasing the level of demand facing their products. These pecuniary externalities correspond to "first-order" effects because of monopoly markups: firms do not capture the full gains from increased production, which instead creates first-order gains for households and for other firms that can sell more.

The interpretation for this result suggested by Murphy, Shleifer and Vishny is to consider the equilibrium with no investment in the new technology as representing a "development trap," where the economy remains in "underdevelopment" because no firm undertakes the investment in new technology and this behavior implies that the demand necessary to make such investments profitable is absent. In contrast, the equilibrium with investment in new technology is interpreted as corresponding to "industrialization". According to this interpretation, societies that can somehow *coordinate* on the equilibrium with investment (either because private expectations are aligned or because of some type of government action) will industrialize and realize both economic growth and Pareto improvement. As such, this model is argued to provide a formalization of the "big push" type industrialization described by economists such as Nurske or Rosenstein-Rodan. Although the idea of the big push and the aggregate demand externalities are attractive, the model here suffers from a number of obvious shortcomings. First, even though the process of industrialization is a dynamic one, the model here is static. Therefore, it does not allow a literal interpretation of a society being first in the no investment equilibrium and then changing to the investment equilibrium and industrializing. Second, as already discussed in Chapter 4, models with multiple equilibria do not provide satisfactory theories of development, since it is difficult to imagine a society remaining unable to coordinate on a simple range of actions that would make *all* households (and firms) better-off. Instead, it is much more likely that the ideas related to aggregate demand externalities (or other potential forces leading to multiple equilibria) are more important as sources of persistence or as mechanisms generating multiple steady states (while still maintaining a unique equilibrium path).

21.6. Inequality, Credit Market Imperfections and Human Capital

The previous section illustrated how aggregate demand externalities can generate development traps. Investment by different firms may require coordination, leading to multiple equilibria. Underdevelopment may be thought to correspond to a situation in which the coordination is on the bad equilibrium, and the development process starts with the "big push," ensuring coordination to the high-investment equilibrium. Here I illustrate a related set of issues in the context of the impact of the distribution of income on human capital under imperfect credit markets. In contrast to the previous section, I will emphasize the possibility of multiple steady states (rather than multiple equilibria). In addition, while I focus on human capital investments, inequality and credit market problems influence not only human capital investments, but also business creation, occupational choices and other aspects of the organization of production. Nevertheless, models focusing on the link between inequality and human capital are both more tractable and also constitute a natural continuation of the theories of human capital investments presented in Chapter 10.

21.6.1. A Simple Case With No Borrowing. When credit markets are imperfect, a major determinant of human capital investments will be the distribution of income (as well as the degree of imperfection in credit markets). I start with a discussion of the simplest case in which there is no borrowing or lending, which introduces an extreme form of credit market problems. I then enrich this model by introducing imperfect credit markets, where the cost of borrowing is greater than the interest rate received by households engaged in saving.

The economy consists of continuum 1 of dynasties. Each individual lives for two periods, childhood and adulthood, and begets an offspring in his adulthood. There is consumption only at the end of adulthood. Preferences are given by

$$(1-\delta)\log c_i(t)+\delta\log e_i(t+1)$$

where c is consumption at the end of the individual's life, and e is the educational spending on the offspring of this individual. The budget constraint is

$$c_i(t)+e_i(t+1)\leq w_i(t),$$

where w is the wage income of the individual. Notice that preferences here have the "warm glow" type altruism which we encountered in Chapter 9 and in Section 21.2 above. In particular, parents do not care about the utility of their offspring, but simply about what they bequeath to them, here education. As usual, this significantly simplifies the analysis.

The labor market is competitive, and wage income of each individual is simply a linear function of his human capital, $h_i(t)$:

$$w_i(t)=Ah_i(t)$$

Human capital of the offspring of individual i of generation t in turn is given by

$$h_i(t+1)=\begin{cases} e_i(t)^{\gamma} & \text{if } e_i(t)\geq 1 \\ \bar{h} & \text{if } e_i(t)<1 \end{cases}, \tag{21.52}$$

where $\gamma\in(0,1)$ and $\bar{h}\in(0,1)$ is some minimum level of human capital that the individual will attain even without any educational spending. Once spending exceeds a certain level (here set equal to 1), the individual starts benefiting from the additional spending and accumulates further human capital (though with diminishing returns since $\gamma<1$).

This equation introduces a crucial feature necessary for models of credit market imperfections to generate multiple equilibria or multiple steady states; a *non-convexity* in the technology of human capital accumulation. Exercise 21.9 shows that this non-convexity plays a crucial role in the results of this subsection.

Given this description, the equilibrium is straightforward to characterize. Each individual will choose the spending on education that maximizes his own utility. This immediately implies the following "saving rate" in terms of education

$$e_i(t)=\delta w_i(t)=\delta Ah_i(t). \tag{21.53}$$

This rule has one unappealing feature (not crucial for any of the results): because parents derive utility from educational spending on their children, they will invest in education even when $e_i(t)<1$; but in this case, educational spendings are wasted (they do not translate into higher human capital of the offspring).

To obtain stark results, let us also assume that

$$\delta A>1>\delta A\bar{h}. \tag{21.54}$$

Now, let us look at the dynamics of human capital for a particular dynasty i. If at time 0, $h_i(0)<(\delta A)^{-1}$, then (21.53) implies that $e_i(t)<1$, so the offspring will have $h_i(1)=\bar{h}$.

Given (21.54), $h_i(1) = \bar{h} < (\delta A)^{-1}$, and repeating this argument, $h_i(t) = \bar{h} < (\delta A)^{-1}$ for all t. Therefore, a dynasty that starts with $h_i(0) < (\delta A)^{-1}$ will never reach a human capital level greater than $\bar{h}$.

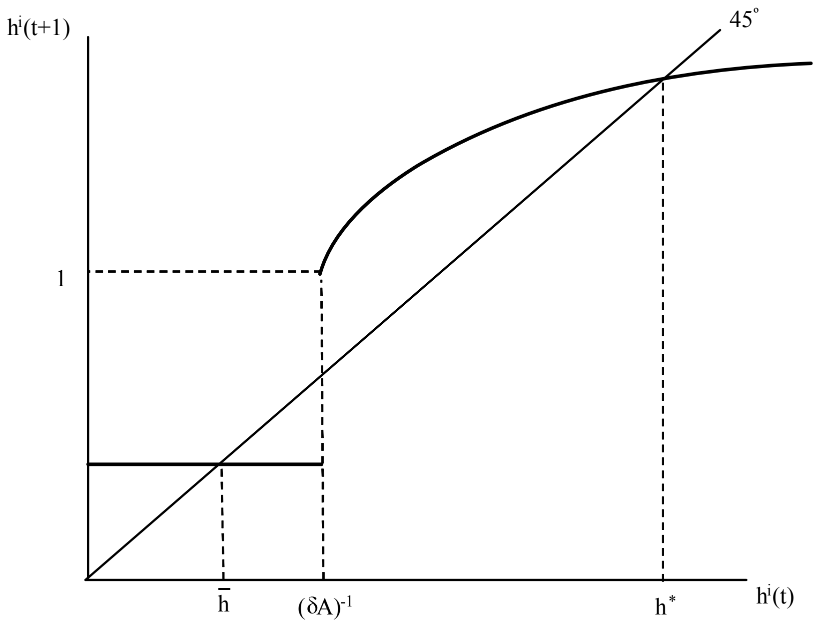

Figure 21.8. Dynamics of human capital with nonconvexities and no borrowing.

Next consider a dynasty with $h_i(0) > (\delta A)^{-1}$. Then, from (21.54), $h_i(1) = (\delta A h_i(0))^{\gamma} > 1$, so this dynasty will gradually accumulate more and more human capital over generations and ultimately reach the "steady state" given by $h^* = (\delta A h^*)^{\gamma}$ or

$$h^* = (\delta A)^{\frac{\gamma}{1-\gamma}} > 1.$$

Naturally, this description applies to a dynasty with $h_i(0) \in ((\delta A)^{-1}, h^*)$. If $h_i(0) > h^*$, then the dynasty would have too much human capital and would decumulate human capital.

Figure 21.8 illustrates the dynamics of individual human capital decisions. It shows that there are two steady-state levels of human capital for individuals, $\bar{h}$ and $h^* > \bar{h}$. An important question when there are multiple steady states is where, given initial conditions, the economy (or a particular individual) will converge to. Assume for now that, even though there are multiple steady states, the equilibrium is unique (meaning that given initial conditions there is a unique equilibrium path—this will be the case in all the models discussed in this section). Then, equilibrium dynamics are represented by a dynamical system with the only difference that there are multiple steady states. Each (locally) asymptotically stable steady state will have a *basin of attraction*, meaning a set of initial conditions, which will ultimately lead to this particular steady state. Both steady states in the model studied here are asymptotically stable and Figure 21.8 plots their basins of attractions. In particular, inspection of this figure shows that dynasties with $h_i(0) < (\delta A)^{-1}$ will tend to the lower steady state level of human capital, $\bar{h}$, while those with $h_i(0) > (\delta A)^{-1}$ will tend to the higher level, h^*.

This figure also reveals why the dynamics in this model are so simple; the dynamics of the human capital of a single individual contain all the information relevant for the dynamics of the human capital and income of the entire economy. This is because there are no prices (such as the rate of return to human capital or the interest rate) that are determined in equilibrium here. For this reason, dynamics in this type of models are sometimes described as *Markovian*—because they are summarized by the Markov process describing the behavior of the human capital of a single individual (without any general equilibrium interactions). Markovian models are much more tractable than those where dynamics of inequality depend on equilibrium prices. An example of this richer type of model is given in Exercise 21.13.

The most important implication of this analysis is again the presence of poverty traps. This is most clearly illustrated by an economy with two groups starting at income levels h_1 and $h_2 > h_1$ such that $(\delta A)^{-1} < h_2$. Now if inequality (poverty) is high so that $h_1 < (\delta A)^{-1}$, a significant fraction of the population will never accumulate much human capital. In contrast, if inequality is limited so that $h_1 > (\delta A)^{-1}$, all agents will accumulate human capital, eventually reaching h^*. This example also illustrates that there are (many) multiple steady states in this economy. Depending on the fraction of dynasties that start with initial human capital below $(\delta A)^{-1}$, any fraction of the population may end up at the low level of human capital, $\bar{h}$. The greater is this fraction, the poorer is the economy.

There are certain parallels between the multiplicity of steady states here and the multiple equilibria highlighted in the model of the previous section. Nevertheless, the differences are in fact more important. In the model of the previous section, there are multiple equilibria in a static model. Thus nothing determines which equilibrium the economy will be in. At best, we can appeal to "*expectations*," arguing that the better equilibrium will emerge when everybody expects the better equilibrium to emerge. One can informally appeal to the role of "*history,*" for example, suggesting that if an economy has been in the low investment equilibrium for a while, it is likely to stay there, but this argument is misleading. First of all, the model is a static one, thus a discussion of an economy "that has been in the low equilibrium for a while" is not quite meaningful. Secondly, even if the model were turned into a dynamic one by repeating it over time, the history of being in one equilibrium for a number of periods will have no effect on the existence of multiple equilibria at the next period. In particular, each static equilibrium would still remain an equilibrium in the "dynamic" environment, and the economy could suddenly jump from one equilibrium to another. This highlights that models with multiple equilibria have a degree of indeterminacy that is both theoretically awkward and empirically difficult to map to reality. Instead, models with multiple steady states avoid these thorny issues. The equilibrium is *unique*, but the initial conditions determine where the dynamical system will eventually end up. Because the equilibrium is unique, there is no issue of indeterminacy or expectations affecting the path of the economy. But also, because multiple steady states are possible, the model can be useful for thinking about potential development traps.

This model also shows the importance of the distribution of income in an economy with imperfect credit markets (here with no credit markets). In particular, the distribution of income affects which individuals will be unable to invest in human capital accumulation and thus influences the long-run income level of the economy. For this reason, models of this sort are sometimes interpreted as implying that an unequal distribution of income will lead to lower output (and growth). The above example with two classes seems to support this conclusion. However, this is not a general result and it is important to emphasize that this class of models does not make specific predictions about the relationship between inequality and growth. To illustrate this, consider again the same economy with two classes, but now starting with $h_1 < h_2 < (\delta A)^{-1}$. In this case, neither group will accumulate human capital, but redistributing resources away from group 1 to group 2 (thus increasing inequality), so that we push group 2 to $h_2 > (\delta A)^{-1}$, would increase human capital accumulation. This is a general feature: in models with non-convexities, there are no unambiguous results about whether greater inequality is good or bad for economic growth; it depends on whether greater inequality pushes more people below or above the critical thresholds. Somewhat sharper results can be obtained about the effect of inequality on human capital accumulation and development under additional assumptions. Exercise 21.10 presents a parameterization of inequality in the model here, which delivers the results that greater inequality leads to lower human capital and lower output per capita in relatively rich economies, but to greater investments in human capital in poorer economies.

21.6.2. Human Capital Investments with Imperfect Credit Markets. I now enrich the environment in the previous subsection by introducing credit markets following Galor and Zeira's (1993) model. Each individual still lives for two periods. In his youth, he can either work or acquire education. The utility function of each individual is

$$(1-\delta)\log c_i(t) + \delta \log b_i(t),$$

where again c denotes consumption at the end of the life of the individual. The budget constraint is

$$c_i(t) + b_i(t) \leq y_i(t),$$

where $y_i(t)$ is individual i's income at time t. Note that preferences still take the "warm glow" form, but the utility of the parent now depends on monetary bequest to the offspring rather than the level of education expenditures. It will now be the individuals themselves who will use the monetary bequests to invest in education. Also, the logarithmic formulation once again ensures a constant saving rate equal to δ.

Education is a binary outcome, and educated (skilled) workers earn wage w_s while uneducated workers earn w_u. The required education expenditure to become skilled is h, and workers acquiring education do not earn the unskilled wage, w_u, during the first period of their lives. The fact that education is binary decision introduces the aforementioned non-convexity in human capital investment decisions.[3]

Imperfect capital markets are modeled by assuming that there is some amount of monitoring required for loans to be paid back. The cost of monitoring creates a wedge between the borrowing and the lending rates. In particular, assume that there is a linear savings technology open to all agents, which fixes the lending rate at some constant r. However, the borrowing rate is $i > r$, because of costs of monitoring necessary to induce agents to pay back the loans (see Exercise 21.12 for a more micro-founded version of these borrowing costs).

Also assume that

$$w_s - (1+r)\,h > w_u\,(2+r) \tag{21.55}$$

which implies that investment in human capital is profitable when financed at the lending rate r.

Consider an individual with wealth x. If $x \geq h$, (21.55) implies that the individual will invest in education. If $x < h$, then whether it is profitable to invest in education will depend on the wealth of the individual and on the borrowing interest rate, i.

Let us now write the utility of this individual (with $x < h$) in the two scenarios, and also the bequest that he will leave to his offspring. These are

$$\begin{aligned} U_s\,(x) &= \log\,(w_s + (1+i)\,(x-h)) + \log\,(1-\delta)^{1-\delta}\,\delta^{\delta} \\ b_s\,(x) &= \delta\,(w_s + (1+i)\,(x-h))\,, \end{aligned}$$

when he invests in education. When he chooses not to invest, then

$$\begin{aligned} U_u\,(x) &= \log\,((1+r)\,(w_u + x) + w_u) + \log\,(1-\delta)^{1-\delta}\,\delta^{\delta} \\ b_u\,(x) &= \delta\,((1+r)\,(w_u + x) + w_u)\,. \end{aligned}$$

Comparing these expressions, it is clear that an individual prefers to invest in education if and only if

$$x \geq f \equiv \frac{(2+r)\,w_u + (1+i)\,h - w_s}{i-r}$$

The dynamics of individual wealth can then be obtained simply by using the bequests of unconstrained, constrained-investing and constrained-non-investing agents.

More specifically, the equilibrium correspondence describing equilibrium dynamics is

$$x\,(t+1) = \begin{cases} b_u\,(x\,(t)) = \delta\,((1+r)\,(w_u + x\,(t)) + w_u) & \text{if } x\,(t) < f \\ b_s\,(x\,(t)) = \delta\,(w_s + (1+i)\,(x\,(t) - h)) & \text{if } h > x\,(t) \geq f \\ b_n\,(x\,(t)) = \delta\,(w_s + (1+r)\,(x\,(t) - h)) & \text{if } x\,(t) \geq h \end{cases} \tag{21.56}$$

Equilibrium dynamics can now be analyzed diagrammatically by looking at the graph of (21.56), which is shown in Figure 21.9. As emphasized in the context of the model of the previous subsection, the curve corresponding to equation (21.56) describes both the behavior

[3] An alternative to nonconvexities in human capital investments is presented in Galor and Moav (2004), who show that multiple steady states are possible when there are no nonconvexities, credit markets are imperfect and the marginal propensity to save is higher for richer dynasties. This assumption is motivated by Kaldor's (1957) paper and was discussed in Exercise 2.12 in Chapter 2.

of the wealth of each individual and the behavior of the wealth distribution in the aggregate economy. This is again a feature of the "Markovian" nature of the current model.

Now define x^* as the intersection of the equilibrium curve (21.56) with the 45 degree line when the equilibrium correspondence is steeper. Such an intersection will exist when the borrowing interest rate, i, is large enough. Suppose this is the case. Then, Figure 21.9 makes it clear that there will be three intersections between (21.56) and the 45 degree line, $\bar{x}_U$, x^* and $\bar{x}_S$. Moreover, the figure shows that x^* corresponds to an unstable steady state, while the other two are locally asymptotically stable.

FIGURE 21.9. Multiple steady-state equilibria in the Galor and Zeira model.

The basis of attraction of the steady states for $\bar{x}_U$ and $\bar{x}_S$ are also easy to obtain from this figure. In particular, all individuals with $x(t) < x^*$ converge to the wealth level $\bar{x}_U$, while all those with $x(t) > x^*$ converge to the greater wealth level $\bar{x}_S$. Thus the basin of attraction of $\bar{x}_U$ is $[0, x^*]$, and this corresponds to a "poverty trap," in the sense that individuals (dynasties) with initial wealth in this interval will converge to $\bar{x}_U$. The initial distribution of income again has a first-order effect on the efficiency and income level of the economy. If the majority of the individuals start with $x(t) < x^*$, the economy will have low productivity, low human capital and low wealth. Therefore, this model extends the insights of the simple model with no borrowing from the previous subsection to a richer environment in which individuals make forward-looking human capital investments. The key is again the interaction between credit market imperfections (which here make the interest rate for borrowing greater than the interest rate for saving) and inequality. As in the model of the previous subsection, it is straightforward to construct examples where an increase in inequality can lead to either worse or better outcomes depending on whether it pushes more individuals into the basin of attraction of the low steady state.

An important feature of the model of this subsection is that because it allows individuals to borrow and lend in financial markets, it enables an investigation of the implications of financial development for human capital investments. In an economy with better financial institutions, the wedge between the borrowing rate and the lending rate will be smaller, that is, i will be smaller for a given level of r. With a smaller i, more agents will escape the poverty trap, and in fact, the poverty trap may not exist at all (there may not be an intersection between (21.56) and the 45 degree line where (21.56) is steeper). This shows that financial development not only improves risk sharing as demonstrated in Section 21.1, but in addition, by relaxing credit market constraints, it contributes to human capital accumulation.

Although the model in this section is considerably richer than that in the previous subsection, it is still a partial equilibrium model. Multiple steady states are possible for different individuals as a function of their initial level of human capital (or wealth), but individual dynamics are not affected by general equilibrium prices. Galor and Zeira (1993), Banerjee and Newman (1994), Aghion and Bolton (1997) and Piketty (1997) consider richer environments in which income dynamics of each dynasty (individual) are affected by general equilibrium prices (such as the interest rate or the wage rate), which are themselves a function of the income inequality. Exercise 21.11 shows that the type of multiple steady states generated by the model presented here may not be robust to the addition of noise in income dynamics—instead of multiple steady states, the long-run equilibrium may generate a stationary distribution of human capital levels, though this stationary distribution would exhibit a large amount of persistence.[4] In contrast, models in which prices are determined in general equilibrium and affect wealth (income) dynamics can generate more "robust" multiplicity of steady states.

21.7. Towards a Unified Theory of Development and Growth?

There has been a unified theme to the models discussed in this chapter. They have either emphasized the transformation of the economy and the society over the process of development or potential reasons for why such a transformation might be halted. This transformation takes the form of the structure of production changing, the process of industrialization getting underway, a greater fraction of the population migrating from rural areas to cities, financial markets becoming more developed, mortality and fertility rates changing via health improvements and the demographic transition, and the extent of inefficiencies and market failures becoming less pronounced over time. In many instances, the driving force for this process is reinforced by the structural transformation that it causes.

[4]The reader will note that this is related to the "Markovian" nature of the model. Markovian models can generate multiple steady states because the Markov chain or the Markov process implied by the model is not ergodic (e.g., poor individuals can never accumulate to become rich). A small amount of noise then ensures that different parts of the distribution "communicate," making the Markov process ergodic, thus removing the multiplicity of steady states.

My purpose in this section is not to offer a unified model of structural transformations and market failures in development. An attempt to pack many different aspects of development into a single model often leads to a framework that is complicated, whereas I believe that relatively abstract representations of reality are more insightful. Moreover, the literature has not made sufficient progress for us to be able to develop a unified framework. Instead, I wish to provide a reduced-form model intended to bring out the salient common features of the models presented in this chapter.

In all of the models presented in this and the previous chapters, economic development is associated with capital deepening, that is, with greater use of capital instead of human labor. Thus we can also approximate the growth process with an increase in the capital-labor ratio of the economy, $k(t)$. This does not necessarily mean that capital accumulation is the engine of economic growth. In fact, previous chapters have emphasized how technological change is often at the root of the process of economic growth (and economic development) and thus capital deepening may be the result of technological change. Moreover, Section 21.4 showed how the crucial variable capturing the stage of development might be the distance of an economy's technology to the world technology frontier. Nevertheless, even in these cases, an increase in the capital-labor ratio will take place along the equilibrium path and can thus be used as a proxy for the stage of development (though in this case one must be careful not to confuse increasing the capital-labor ratio with ensuring economic development). With this caveat in mind, in this section I take the capital-labor ratio as the proxy for the stage of development and for analytical convenience, I use the Solow model to represent the dynamics of the capital-labor ratio.

In particular, consider a continuous-time economy, with output per capita given by

$$y(t) = f(k(t), x(t)), \tag{21.57}$$

where $k(t)$ is the capital-labor ratio and $x(t)$ is some "social variable," such as financial development, urbanization, structure of production, the structure of the family and so on. As usual, f is assumed to be differentiable and also increasing and concave in k. Moreover, the social variable x potentially affects the efficiency of the production process and thus is part of the per capita production function in (21.57). As a convention, suppose that an increase in x corresponds to "structural change" (for example, a move from the countryside to cities). Therefore, f is also increasing in x and the partial derivative with respect to x is nonnegative, that is, $f_x \geq 0$. Naturally, not all structural change is beneficial. Nevertheless, for simplicity, I focus on the case in which f is increasing in x.

Suppose that structural change can be represented by the differential equation

$$\dot{x}(t) = g(k(t), x(t)), \tag{21.58}$$

where g is also assumed to be twice differentiable. Since x corresponds to structural change associated with development, g should be increasing in k, and in particular, its partial derivative with respect to k is strictly positive, that is, $g_k > 0$. Moreover, standard mean reversion type reasoning suggests that the case in which the derivative g_x is negative is the most reasonable benchmark. If x is above its "natural level," it should decline and if it is below its natural level, it should increase. Motivated by this, let us also assume that $g_x < 0$.

Capital accumulates according to the Solow growth model is in Chapter 2, so that

$$\dot{k}(t) = sf(k(t), x(t)) - \delta k(t), \tag{21.59}$$

where I have suppressed population growth and there is no technological change for simplicity. For a fixed x, capital naturally accumulates in an identical fashion to that in the basic Solow model. The structure of this economy is slightly more involved because $x(t)$ also changes.

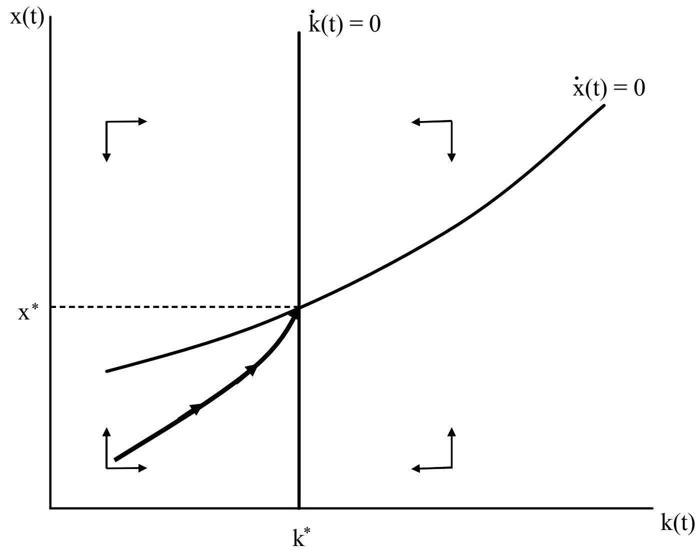

Figure 21.10. Capital accumulation and structural transformation without any effect of the "social variable" x on productivity.

First consider the case in which $f_x(k, x) \equiv 0$ so that the social variable x has no effect on productivity. Dynamics in this case are shown in Figure 21.10. The thick vertical line corresponds to the locus for $\dot{k}(t)/k(t) = 0$—it represents the steady state of the differential equation (21.59). This locus takes the form of a vertical line, since only a single value of $k(t)$, k^*, is consistent with steady state. The upward sloping line, on the other hand, corresponds to (21.58) and shows the locus of the values of k and x such that $\dot{x}(t)/x(t) = 0$. It is upward sloping, since g is increasing in k and decreasing in x. The laws of motion represented by the arrows follow straightforwardly from (21.58) and (21.59). For example, when $k(t) < k^*$, (21.59) implies that $k(t)$ will increase. Similarly, when $x(t)$ is above the $\dot{x}(t)/x(t) = 0$ locus, (21.58) implies that $x(t)$ will decrease. Given these laws of motion, it is straightforward to see that the dynamical system representing the equilibrium of this model is globally stable and starting with any $k(0) > 0$ and $x(0) > 0$, the economy will travel towards the unique steady state (k^*, x^*). Now consider the dynamics of a less-developed economy, that is, an economy that starts with a low level of capital-labor ratio, $k(0)$, and a low level of the social variable,

Figure 21.11. Capital accumulation and structural transformation with multiple steady states.

$x(0)$. Then, development in this economy will take place with gradual capital deepening and a corresponding increase in $x(t)$ towards x^*, which can be viewed as a reduced-form representation of development-induced structural change.

Next, consider the more interesting case in which $f_x(k,x) > 0$. In this case, the locus for $\dot{k}(t)/k(t) = 0$ will also be upward sloping, since $f_x > 0$ and the right-hand side of (21.59) is decreasing in k by the standard arguments (in particular, because of the fact that by the strict concavity of $f(k,x)$ in k, $f(k,x)/k > f_k(k,x)$ for all k and x, see Exercise 21.14). A steady state is again given by the intersection of the loci for $\dot{k}(t)/k(t) = 0$ and $\dot{x}(t)/x(t) = 0$. Since both of these are now upward sloping, multiple steady states are possible as shown in Figure 21.11. These multiple steady states may correspond to the multiple equilibria arising from aggregate demand externalities or from the presence of imperfect credit markets. The low steady state (k', x') corresponds to a situation in which the social variable x is low and this depresses productivity, making the economy settle into an equilibrium with a low capital-labor ratio. In contrast in the high steady state (k^*, x^*), the high level of x supports greater productivity and thus a greater capital-labor ratio consistent with steady state. Moreover, it can be verified that both the low and the high steady states are typically locally stable, so that starting from the neighborhood of one, the economy will converge to the "nearest" steady state and will tend to stay there. This highlights the importance of *historical factors* in the development process. If historical factors or endowments placed the economy in the basin of attraction of the low steady state, the economy will converge to this steady state corresponding to a "development trap". Interestingly, this development trap is, at least in part, caused by lack of structural change (that is, by a low value of the social variable x).

Figure 21.11 makes it clear that such multiplicity requires the locus for $\dot{k}(t)/k(t) = 0$ to be relatively flat, at least over some range. Inspection of (21.59) shows that this will be the case when $f_x(k, x)$ is large, at least over some range. Intuitively, multiple steady states can only arise when the social variable x (or structural change) has a large effect on productivity.

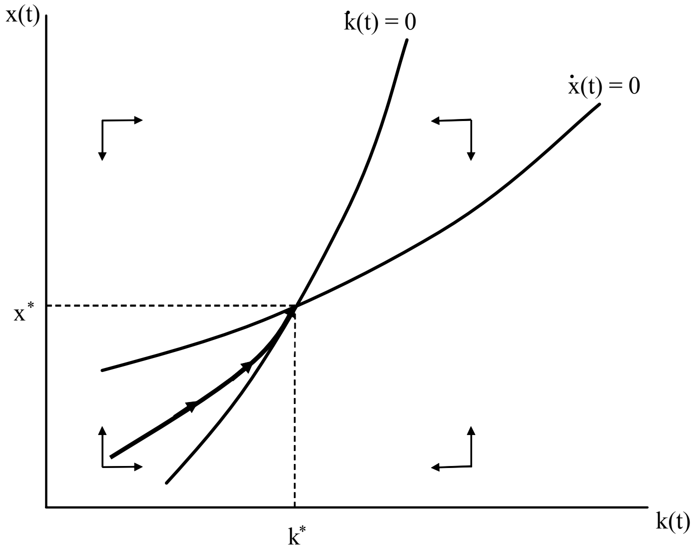

Figure 21.12. Capital accumulation and structural transformation when the "social variable" x affects but there exists a unique steady state.

Perhaps more interesting than multiple steady states is the situation in which the same forces are present, but a unique steady state exists. The same reasoning suggests that this will be the case when $f_x(k, x)$ is relatively small. In this case, the locus for $\dot{k}(t)/k(t) = 0$ will be everywhere steeper than the locus for $\dot{x}(t)/x(t) = 0$. This case is plotted in Figure 21.12 and the unique steady state is given by (k^*, x^*). The laws of motion represented by the arrows again follow from the inspection of the differential equations (21.58) and (21.59). This figure shows that the unique steady state is globally stable (see Exercise 21.14 for a formal proof). Consider, once again, a less-developed economy starting with a low level of capital-labor ratio, $k(0)$, and a low level of the social variable, $x(0)$. The dynamics in this case are qualitatively similar to those in Figure 21.10. However, the economics is slightly different. Capital accumulation (capital deepening) leads to an increase in $x(t)$ as before, but now this structural change also improves productivity as in the models in Section 17.6 in Chapter 17 and in Sections 21.3 and 21.1 of this chapter. This increase in productivity leads to faster capital accumulation and there is a *self-reinforcing* ("cumulative") process of development, with economic growth leading to structural changes facilitating further growth.

However, since the effect of x on productivity is limited, this process ultimately takes us towards a unique steady state.

This reduced-form representation of structural change, therefore, captures some of the salient features emphasized in this chapter. It is not meant to be a unified model; on the contrary, rather than combining multiple dimensions of structural change, it presents an abstract representation emphasizing how the process of development, corresponding to capital accumulation, can go hand-in-hand with structural change, which may in turn increase productivity and facilitate further capital accumulation. Developing a truly unified model of economic development and structural change is an area for future work.

21.8. Taking Stock

This chapter provided a large number of models focusing on various aspects of the structural transformation accompanying economic development. As emphasized in the previous section, there is no single framework unifying all these distinct aspects, even though there are many common themes in these models. The previous section was an attempt to bring out these common themes. Instead of repeating these commonalities, I would like to conclude by pointing out that many of the topics covered in this chapter are at the frontier of current research and much still remains to be done. Economic development is intimately linked to economic growth, but it may require different, even specialized, models that do not just focus on balanced growth and the orderly growth behavior captured by the neoclassical and endogenous technology models. These models may also need to take market failures and how these market failures might change over time more seriously. This view stems from the recognition that the essence of economic development is the process of structural transformation, including financial development, the demographic transition, migration, urbanization, organizational change and other social changes.

Another potentially important aspect of economic development is the possibility that the inefficiencies in the organization of production, credit markets and product markets may culminate in potential development traps. These inefficiencies may stem from lack of coordination in the presence of aggregate demand externalities or from the interaction between imperfect credit markets and human capital investments. These topics not only highlight some of the questions that need to be addressed for understanding the process of economic development, but they also bring a range of issues that are often secondary in the standard growth literature to the forefront of analysis. These include, among other things, the organization of financial markets, the distribution of income and wealth and issues of incentives, such as problems of moral hazard, adverse selection and incomplete contracts both in credit markets and in production relationships.

The recognition that the analysis of economic development necessitates a special focus on these topics also opens the way for a more constructive interaction between empirical development studies and the theories of economic development surveyed in this chapter. As already noted above, there is now a large literature on empirical development economics, documenting the extent of credit market imperfections, the impact of inequality on human capital investments and occupation choices, the process of social change and various other market failures in less-developed economies. By and large, this literature is about market failures in less-developed economies and sometimes also focuses on how these market failures can be rectified. The standard models of economic growth do not feature these market failures. A fruitful area for future research is then the combination of theoretical models of economic growth and development (that pay attention to market failures) with the rich empirical evidence on the incidence, characterization and costs of these market failures. This combination will have the advantage of being theoretically rigorous and empirically grounded, and perhaps most importantly, it can focus on what I believe to be the essence of development economics—the questions of why some countries are less developed, how they can grow more rapidly and how they can jumpstart the process of structural transformation necessary for economic development.

21.9. References and Literature

By its nature, this chapter has covered a large amount of material. My selection of topics has reflected my own interests and was also motivated by my desire to keep this chapter from becoming even longer than it already is.

Section 21.1 scratches the surface of a rich literature on financial development and economic growth. On the theoretical side, Townsend (1979), Greenwood and Jovanovic (1990) and Bencivenga and Smith (1991) focus on the interaction between financial development on the one hand and risk sharing and the allocation of funds across different tasks and individuals on the other. Obstfeld (1994) and Acemoglu and Zilibotti (1997) focus on the relationship between financial development and the diversification of risks. There is also a large empirical literature looking at the effect of financial development on economic growth. An excellent survey of this literature is provided in Levine (2005). Some of the most well-known empirical papers include King and Levine (1993), which documents the cross-country correlation between measures of financial development and economic growth, Rajan and Zingales (1998), which shows that lack of financial development has particularly negative effects on sectors that have greater external borrowing needs, and Jayaratne and Strahan (1996), which documents how banking deregulation that increased competition in US financial markets led to more rapid financial and economic growth within the United States. In discussing financial development, I also mentioned the literature on the Kuznets curve. There is no consensus on whether there is a Kuznets curve. Work that focuses on historical data, such as Lindert and Williamson (1976) or Bourguignon and Morrison (2002), reports aggregate patterns consistent with a Kuznets curve, while studies using panels of countries in the postwar era, such as Fields (1980), do not find a consistent pattern resembling the Kuznets curve.

The literature on fertility, the demographic transition and growth is also vast. The main trends in world population and cross-country differences in population growth are summarized in Livi-Bacci (1997) and Maddison (2003). The idea that parents face a tradeoff between the numbers and the human capital of their children—the quality and quantity tradeoff—was proposed by Becker (1981). The aggregate patterns in Livi-Bacci (1997) are consistent with this idea, though there is little micro evidence supporting this tradeoff. Recent work on microdata, by Black, Devereux and Salvanes (2005), Angrist, Lavy and Schlosser (2006) and Qian (2007), looks at evidence from Norway, Israel and China, but does not find strong support for the quality-quantity tradeoff. Fertility choices were first introduced into growth models by Becker and Barro (1988) and Barro and Becker (1989). Becker, Murphy and Tamura (1990) provide the first endogenous growth model with fertility choice. More recent work on the demographic transition and the transition from a Malthusian regime to one of sustained growth include Goodfriend and McDermott (1995), Galor and Weil (1996, 2000), Hansen and Prescott (2002), and Doepke (2004). Kalemli-Ozcan (2002) and Fernandez-Villaverde (2003) focus on the effect of declining mortality on fertility choices in a growth context. A recent series of papers by Galor and Moav (2002, 2004) combine fertility choice, quality-quantity tradeoff and natural selection. Galor (2005) provides an excellent overview of this literature. The first model presented in Section 21.2 is a simplified version of Malthus's classic model in his (1798) book, while the second model is a simplified version of Becker and Barro (1988) and Galor and Weil (1999).

Urbanization is another major aspect of the process of economic development. Bairoch (1988) provides an overview of the history of urbanization. The first model in Section 21.3 builds on Arthur Lewis's (1954) classic, which argued that early development can be viewed as a situation in which there is surplus labor available to the modern sector, thus growth is constrained by capital and technology but not by labor. Harris and Todaro's well-known (1970) paper also emphasizes the importance of model of migration, though it features free migration between rural and urban areas and suggests that unemployment in urban areas will be the key equilibriating variable.

The second model, presented in subsection 21.3.2, is inspired by Banerjee and Newman (1998) and Acemoglu and Zilibotti (1999). Banerjee and Newman emphasize the advantage of smaller rural communities in reducing moral hazard problems in credit relations and show how this interacts with the process of urbanization, which involves individuals migrating to areas where their marginal product is higher. Acemoglu and Zilibotti argue that development leads to "information accumulation". In particular, as more individuals perform similar tasks, more socially useful information is revealed and this enables more complex contractual and production relations. Subsection 21.3.2 also touched upon another important aspect of social and economic relations in less-developed economies, the importance of community enforcement. Clifford Geertz (1963) emphasizes the importance of community enforcement mechanisms and how they may sometimes conflict with markets.

Section 21.4 builds on Acemoglu, Aghion and Zilibotti (2006).

Section 21.5 is based on Murphy, Shleifer and Vishny's famous (1989) paper, which formalized ideas first proposed by Rosenstein-Rodan (1943). Other models that demonstrate the possibility of multiple equilibria in monopolistic competition models featuring non-convexities include Kiyotaki (1988), who derives a similar result in a model with endogenous labor supply choices as well as investment decisions. Matsuyama (1995) provides an excellent overview of these models and a clear discussion of why pecuniary externalities can lead to multiple equilibria in the presence of monopolistic competition.

The distinction between multiple equilibria and multiple steady states is discussed in Krugman (1991) and Matsuyama (1991). Both of these papers highlight that in models with multiple equilibria, expectations determine which equilibrium will emerge, while with multiple steady states, there can be (or there often is) a unique equilibrium and initial conditions (history) determine where the economy will end up.

The model in subsection 21.6.2 is based on the first model in Galor and Zeira (1993). Similar ideas are investigated in Banerjee and Newman (1994) in the context of the effect of inequality on occupational choice, and in Aghion and Bolton (1997) and Piketty (1997) in the context of the interaction between inequality and entrepreneurial investments. Other work on dynamics of inequality and its interactions with efficiency include Loury (1981), Tamura (1991), Benabou (1996), Durlauf (1996), Fernandez and Rogerson (1996), Glomm and Ravikumar (1992) and Acemoglu (1997b).

21.10. Exercises

Exercise 21.1. Analyze the equilibrium of the economy in Section 21.1, relaxing the assumption that each individual has to invest either all or none of his wealth in the risky saving technology. Does this generalization affect the qualitative results derived in the text?

Exercise 21.2. Consider the economy in Section 21.1.

(1) Show that in equation (21.5), $K(t+1)$ is everywhere increasing in $K(t)$ and that there exists some $\bar{K}$ such that the capital stock will grow over time when $K(t) > \bar{K}$.
(2) Can there be more than one steady state level of capital stock in this economy? If so, provide an intuition for this type of multiplicity.
(3) Provide sufficient conditions for the steady state level of capital stock, K^*, to be unique. Show that in this case $K(t+1) > K(t)$ whenever $K(t) < K^*$.

Exercise 21.3. In the model of subsection 21.2.1, suppose that the population growth equation takes the form $L(t+1) = \varepsilon(t)(n(t+1)-1)L(t)$ instead of (21.8), where $\varepsilon(t)$ is a random variable that takes one of two values, $1-\bar{\varepsilon}$ or $1+\bar{\varepsilon}$, reflecting random factors affecting population growth. Characterize the stochastic equilibrium. In particular, plot the stochastic correspondence representing the dynamic equilibrium behavior and analyze how shocks affect population growth and income dynamics.

Exercise 21.4. Characterize the full dynamics of migration, urban capital-labor ratio and wages in the model of subsection 21.3.1 (that is, consider the cases in which conditions 1, 2 and 3 in that subsection do not all hold together).

Exercise 21.5. Consider the model of subsection 21.3.2 and suppose that all individuals have time $t = 0$ utility given by the standard CRRA preferences. Taking the equilibrium path in that subsection as given, find a level of community enforcement advantage ξ that would maximize time $t = 0$ utility. What happens if the actual comparative advantage of community enforcement of villages is greater than this level?

Exercise 21.6. Consider the maximization problem given in (21.31).

(1) Explain why this maximization problem characterizes the equilibrium allocation of workers to tasks. What kind of price system will support this allocation?
(2) Derive the first-order conditions given in (21.32).
(3) Provide sufficient conditions such that the solution to this problem involves all skilled workers being employed at technology $\bar{h}$.
(4) Provide an example in which no worker will be employed at technology $\bar{h}$ even though $A_{\bar{h}} > A_h$ for all $h \in [0, \bar{h}]$.
(5) Can there be a solution where more than two technologies are being used in equilibrium? If so, explain the conditions for such an equilibrium to arise.

Exercise 21.7. Consider a variant of the model in Section 21.4, where firms have an organizational form decision, in particular, they decide whether or not to vertically integrate. For this purpose, consider a slight modification of equation (21.38) where

$$A(\nu, t) = \eta \bar{A}(t-1) + \gamma(\nu, t) A(t-1),$$

with $\gamma(\nu, t) = \underline{\gamma} + \theta(\nu, t)$. Suppose that entrepreneurial effort increases $\theta(\nu, t)$, and the internal organization of the firm affects how much effort the entrepreneur devotes to innovation activities. In particular, suppose that $\theta(\nu, t) = 0$ if there is vertical integration, because the entrepreneur is overloaded and has limited time for innovation activities. In contrast, with outsourcing $\theta(\nu, t) = \theta > 0$. However, when there is outsourcing, the entrepreneur has to share a fraction $\beta > 0$ of the profits with the manager (owner) of the firm to which certain tasks have been outsourced (whereas in a vertically integrated structure, he can keep the entire revenue).

(1) Determine the profit-maximizing outsourcing decision for an entrepreneur as a function of $a(t)$. In particular, show that there exists a threshold $\bar{a}$ such that there will be vertical integration for all $a(t) \leq \bar{a}$ and outsourcing for all $a(t) > \bar{a}$.
(2) Contrast this equilibrium behavior with the growth-maximizing internal organization of the firm.

Exercise 21.8. Show that when multiple equilibria exist in the model of Section 21.5, the equilibrium with investment Pareto dominates the one without.

Exercise 21.9. Consider the model of subsection 21.6.1 and remove the non-convexity in the accumulation equation, (21.52), so that the human capital of the offspring of individual i is given by $h_i(t+1) = e_i(t)^\gamma$ for any level of $e_i(t)$ and $\gamma \in (0, 1)$. Show that there exists a unique level of human capital to which each dynasty will converge . Based on this result, explain the role of non-convexities in generating multiple steady states.

Exercise 21.10. Consider the model of subsection 21.6.1 and suppose that initial inequality is given by a uniform distribution with mean human capital of $h(0)$ and support over $[h(0) - \lambda, h(0) + \lambda]$. Clearly an increase in λ corresponds to greater inequality.

(1) Show that when $h(0)$ is sufficiently small, an increase in λ will increase long-run average human capital and income, whereas when $h(0)$ is sufficiently large, an increase in λ will reduce them. [Hint: use Figures 21.8-21.9].
(2) What other types of distributions (besides uniform) would lead to the same result?
(3) Show that the same result generalizes to the model of 21.6.2.
(4) On the basis of this result, discuss whether we should expect greater inequality to lead to higher income in poor societies and lower income in rich societies. (If your answer is no, then sketch an environment in which this will not be the case).

Exercise 21.11. Consider the model presented in subsection 21.6.2. Make the following two modi☐cations. First, the utility function is now

$$(1-\delta)^{-(1-\delta)}\,\delta^{-\delta}c^{1-\delta}b^{\delta} \tag{21.60}$$

and second, unskilled agents receive a wage of $w_u+\varepsilon$ where ε is a mean-zero random shock.

(1) Suppose that ε is distributed with support $[-\lambda,\lambda]$, and show that if λ is sufficiently close to 0, then the multiple steady states characterized in 21.6.2 "survive" in the sense that depending on their initial conditions some dynasties become high skilled and others become low skilled.
(2) Why was it convenient to change the utility function from the log form used in the text to (21.60)?
(3) Now suppose that ε is distributed with support $[-\lambda,\infty)$, where $\lambda\leq w_u$. Show that in this case there is a unique ergodic distribution of wealth and no poverty trap. Explain why the results here are different from those in part 1?
(4) How would the results be different if, in addition, the skilled wage is equal to $w_s+\upsilon$, where υ is another mean-zero random shock? [Hint: simply sketch the analysis and the structure of the equilibrium without repeating the full analysis of part 3].

Exercise 21.12. (1) In the model of subsection 21.6.2, suppose that each individual can run away without paying his debts, and if he does so, he will never be caught. However, a bank can prevent this by paying a monitoring cost per unit of borrowing equal to m. Suppose that there are many banks competing a la Bertrand for lending opportunities. Under these assumptions, show that all bank lending will be accompanied with monitoring, and the lending rate will satisfy $i=r+m$. Show that in this case all of the results in the text apply.
(2) Next suppose that the bank can prevent individual from running away by paying a ☐xed monitoring cost of M. Under the same assumptions as in part 1 above show that in this case the interest rate charged to an individual that borrows an amount $x-h$ will be $i(x-h)=r+M/(x-h)$. Given this assumption, characterize the equilibrium of the model in subsection 21.6.2. How do the conclusions change in this case?
(3) Next suppose that there is no way of preventing running away by individuals, but if an individual runs away, he will be caught with probability p, and in this case,

a fraction $\lambda \in (0,1)$ of his income will be confiscated. Given this assumption, characterize equilibrium dynamics of the model in subsection 21.6.2. How do the conclusions change in this case?

(4) Now consider an increase in w_s (for a given level of w_u) so that the skill premium in the economy increases. In which of the three scenarios outlined above will this have the largest effect on human capital investments?

EXERCISE 21.13. In this exercise, you are asked to study Banerjee and Newman's (1994) model of occupational choice. The utility of each individual is again $(1-\delta)^{-(1-\delta)}\delta^{-\delta}c^{1-\delta}b^{\delta} - z$, where z denotes whether the individual is exerting effort, with cost of effort normalized to 1. Each agent chooses one of four possible occupations. These are (1) subsistence and no work, which leads to no labor income and has a rate of return on assets equal to $\hat{r} < 1/\delta$; (2) work for a wage v; (3) self-employment, which requires investment I plus the labor of the individual; and (4) entrepreneurship, which requires investment μI plus the employment of μ workers, and the individual himself becomes the boss, monitoring the workers (and does not take part in production). All occupations other than subsistence involve effort. Let us assume that both entrepreneurship and self-employment generate a rate of return greater than subsistence (that is, the mean return for both activities is $\bar{r} > \hat{r}$).

(1) Derive the indirect utility function associated with the preferences above. Show that no individual will work as a worker for a wage less than 1.
(2) Assume that $\mu[I(\bar{r}-\hat{r})-1]-1 > I(\bar{r}-\hat{r})-1 > 0$. Interpret this assumption. [Hint: it relates the profitabilities of entrepreneurship and self-employment at the minimum possible wage of 1].
(3) Suppose that only agents that have wealth $w \geq w^*$ can borrow enough to become self-employed and only agents that have wealth $w \geq w^{**} > w^*$ can borrow μI to become an entrepreneur. Provide intuition for these borrowing constraints.
(4) Now compute the expected indirect utility from the four occupations. Show that if $v > \bar{v} \equiv (\mu-1)(\bar{r}-\hat{r}) I/\mu$, then self-employment is preferred to entrepreneurship.
(5) Suppose the wealth distribution at time t is given by $G_t(w)$. On the basis of the results in part 4, show that the demand for labor in this economy is given by

$$\begin{array}{cl} x = 0 & \text{if } v > \bar{v} \\ x \in [0, \mu(1-G_t(w^{**}))] & \text{if } v = \bar{v} \\ x = \mu(1-G_t(w^{**})) & \text{if } v < \bar{v}, \end{array}$$

(6) Let $\tilde{v} \equiv (\bar{r}-\hat{r}) I > \bar{v}$. Then, show that the supply of labor is given by

$$\begin{array}{cl} s = 0 & \text{if } v < 1 \\ s \in [0, G_t(w^*)] & \text{if } v = 1 \\ s = G_t(w^*) & \text{if } 1 < v < \tilde{v} \\ s \in [G_t(w^*), 1] & \text{if } v = \tilde{v}, \\ s = 1 & \text{if } v > \tilde{v}. \end{array}$$

(7) Show that if $G_t(w^*) > \mu[1 - G_t(w^{**})]$, there will be an excess supply of labor and the equilibrium wage rate will be $v = 1$. Show that if $G_t(w^*) < \mu[1 - G_t(w^{**})]$, there will be an excess demand for labor and the equilibrium wage rate will be $v = \bar{v}$.

(8) Now derive the individual wealth (bequest) dynamics (for a worker with wealth w) as follows: (1) subsistence and no work: $b(t) = \delta\hat{r}w$; (2) worker: $b(t) = \delta(\hat{r}w + v)$; (3) self-employment: $b(t) = \delta(\bar{r}I + \hat{r}(w - I))$; (4) entrepreneurship: $b(t) = \delta(\bar{r}\mu I + \hat{r}(w - \mu I) - \mu v)$. Explain the intuition for each of these expressions.

(9) Now using these wealth dynamics show that multiple steady states with different wealth distributions and occupational choices are possible. In particular, show that the steady-state wealth level of a worker when the wage rate is v will be $w_w(v) = \delta v/(1 - \delta\hat{r})$, while the steady-state wealth level of a self-employed individual will be $w_{se} = \delta(\bar{r} - \hat{r}) I/(1 - \delta\hat{r})$, and the wealth level of an entrepreneur will be $w_e(v) = \delta(\bar{r}\mu I - \hat{r}\mu I - \mu v)/(1 - \delta\hat{r})$. Now show that when $w_w(v = 1) < w^*$ and $w_e(v = \bar{v}) > w^{**}$, a steady state in which the equilibrium wage rate is equal to $v = 1$ would involve workers not accumulating sufficient wealth to become self-employed, while entrepreneurs accumulate enough wealth to remain entrepreneurs. Explain why this is the case. [Hint: it depends on the equilibrium wage rate].

(10) Given the result in part 9, show that if we start with a wealth distribution such that $\mu(1 - G(w^{**})) < G(w^*)$, the steady state will involve an equilibrium wage $v = 1$ and no self-employment, whereas if we start with $\mu(1 - G(w^{**})) > G(w^*)$, the equilibrium wage would be $v = \bar{v}$ and there will be self-employment. Contrast the level of output in these two steady states.

(11) Is the comparison of the steady states in terms of output in this model plausible? Is it consistent with historical evidence? What are the pros and cons of this model relative to the Galor-Zeira model in subsection 21.6.2?

Exercise 21.14. This exercise asks you to analyze the dynamics of the reduced-form model in Section 21.7 more formally.

(1) Show that when $f_x > 0$, the locus for $\dot{k}/k = 0$ implied by (21.58) is an upward sloping curve.

(2) Consider the differential equations (21.58) and (21.59), and a steady state (k^*, x^*). By linearizing the two differential equations around (k^*, x^*), show that if $f_x(k^*, x^*)$ is sufficiently small, the steady state is locally stable.

(3) Provide a uniform bound on $f_x(k, x)$ so that there exists a unique steady state. Show that when this bound applies, the unique steady state is globally stable.

(4) Construct a parameterized example where there are multiple steady states. Interpret the conditions necessary for this example. Do you find them economically likely?