5 THE ENGINE OF GROWTH

As for the Arts of Delight and Ornament, they are best promoted by the greatest number of emulators. And it is more likely that one ingenious curious man may rather be found among 4 million than among 400 persons. . . .

—William Petty, cited in Simon (1981), p. 158

The neoclassical growth model highlights technological progress as the engine of economic growth, and the previous chapter discussed in broad terms the economics of ideas and technology. In this chapter, we incorporate the insights from the previous chapters to develop an explicit theory of technological progress. The model we develop allows us to explore the engine of economic growth, thus addressing the second main question posed at the beginning of this book. We seek an understanding of why the advanced economies of the world, such as the United States, have grown at something like 2 percent per year for the last century. Where does the technological progress that underlies this growth come from? Why is the growth rate 2 percent per year instead of 1 percent or 10 percent? Can we expect this growth to continue, or is there some limit to economic growth?

Much of the work by economists to address these questions has been labeled *endogenous growth theory* or *new growth theory*. Instead of assuming that growth occurs because of automatic and unmodeled (exogenous) improvements in technology, the theory focuses on understanding the economic forces underlying technological progress. An important contribution of this work is the recognition that technological progress occurs as profit-maximizing firms or inventors seek out newer and better mousetraps. Adam Smith wrote that "it is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest" (Smith 1776 [1981], pp. 26–7). Similarly, it is the possibility of earning a profit that drives firms to develop a computer that can fit in your hand, a soft drink with only a single calorie, or a way to record TV programs and movies to be replayed at your convenience. In this way, improvements in technology, and the process of economic growth itself, are understood as an endogenous outcome of the economy.

The specific theory we will develop in this chapter was constructed by Paul Romer in a series of papers, including a 1990 paper titled "Endogeneous Technological Change."[1]

The Romer model treats technological progress as the addition of new varieties of goods to the menu available to the economy; laptop computers are a new type of good compared to desktop computers, and smartphones are a new good compared to laptops. After we have gone through the Romer model, we present an alternative specification of technology based on improving the quality of existing products: computers today are faster and have greater storage than computers in the past. Developed by Aghion and Howitt (1992) and Grossman and Helpman (1991) originally, this alternative is often referred to as a Schumpeterian growth model, as they were anticipated by the work of Joseph Schumpeter in the late 1930s and early 1940s. What we'll see by the end of the chapter is that the predictions regarding the growth rate of technology in the long run do not depend on how we conceive of technological progress.

THE BASIC ELEMENTS OF THE ROMER MODEL

The Romer model endogenizes technological progress by introducing the search for new ideas by researchers interested in profiting from their inventions. The market structure and economic incentives that

[1] The version of the Romer model that we will present in this chapter is based on Jones (1995a). There is one key difference between the two models, which will be discussed at the appropriate time.

are at the heart of this process will be examined in detail in Section 5.2. First, though, we will outline the basic elements of the model and their implications for economic growth.

The model is designed to explain why and how the advanced countries of the world exhibit sustained growth. In contrast to the neoclassical models in earlier chapters, which could be applied to different countries, the model in this chapter describes the advanced countries of the world as a whole. Technological progress is driven by R&D in the advanced world. In the next chapter we will explore the important process of technology transfer and why different economies have different levels of technology. For the moment, we will concern ourselves with how the world technological frontier is continually pushed outward.

As was the case with the Solow model, there are two main elements in the Romer model of endogenous technological change: an equation describing the production function and a set of equations describing how the inputs for the production function evolve over time. The main equations will be similar to the equations for the Solow model, with one important difference.

The aggregate production function in the Romer model describes how the capital stock, K, and labor, L_Y, combine to produce output, Y, using the stock of ideas, A:

$$Y = K^{\alpha} (AL_Y)^{1-\alpha}, \tag{5.1}$$

where α is a parameter between zero and one. For the moment, we take this production function as given; in Section 5.2, we will discuss in detail the market structure and the microfoundations of the economy that underlie this aggregate production function.

For a given level of technology, A, the production function in equation (5.1) exhibits constant returns to scale in K and L_Y. However, when we recognize that ideas (A) are also an input into production, then there are increasing returns. For example, once Steve Jobs and Steve Wozniak invented the plans for assembling personal computers, those plans (the "idea") did not need to be invented again. To double the production of personal computers, Jobs and Wozniak needed only to double the number of integrated circuits, semiconductors, and so on, and find a larger garage. That is, the production function exhibits constant returns to scale with respect to the capital and labor inputs, and therefore must exhibit increasing returns with respect to all three inputs: if you double capital, labor, *and* the stock of ideas, then you will more than double output. As discussed in Chapter 4, the presence of increasing returns to scale results fundamentally from the nonrivalrous nature of ideas.

The accumulation equations for capital and labor are identical to those for the Solow model. Capital accumulates as people in the economy forgo consumption at some given rate, s_K, and depreciates at the exogenous rate δ:

$$\dot{K} = s_K Y - \delta K.$$

Labor, which is equivalent to the population, grows exponentially at some constant and exogenous rate n:

$$\frac{\dot{L}}{L} = n$$

The key equation that is new relative to the neoclassical model is the equation describing technological progress. In the neoclassical model, the productivity term A grows exogenously at a constant rate. In the Romer model, growth in A is endogenized. How is this accomplished? The answer is with a production function for new ideas: just as more automobile workers can produce more cars, we assume that more researchers can produce more new ideas.

According to the Romer model, $A(t)$ is the stock of knowledge or the number of ideas that have been invented over the course of history up until time t. Then, $\dot{A}$ is the number of new ideas produced at any given point in time. In the simplest version of the model, $\dot{A}$ is equal to the number of people attempting to discover new ideas, L_A, multiplied by the rate at which they discover new ideas, $\bar{\theta}$:

$$\dot{A} = \bar{\theta} L_A \tag{5.2}$$

The rate at which researchers discover new ideas might simply be a constant. On the other hand, one could imagine that it depends on the stock of ideas that have already been invented. For example, perhaps the invention of ideas in the past raises the productivity of researchers in the present. In this case, $\bar{\theta}$ would be an increasing function of A. The discovery of calculus, the invention of the laser, and the development of integrated circuits are examples of ideas that have increased the productivity of later research. On the other hand, perhaps the most obvious ideas are discovered first and subsequent ideas are increasingly difficult to discover. In this case, $\bar{\theta}$ would be a decreasing function of A.

This reasoning suggests modeling the rate at which new ideas are produced as

$$\bar{\theta} = \theta A^{\phi}, \tag{5.3}$$

where θ and ϕ are constants. In this equation, $\phi > 0$ indicates that the productivity of research increases with the stock of ideas that have already been discovered; $\phi < 0$ corresponds to the "fishing-out" case in which the fish become harder to catch over time. Finally, $\phi = 0$ indicates that the tendency for the most obvious ideas to be discovered first exactly offsets the fact that old ideas may facilitate the discovery of new ideas—that is, the productivity of research is independent of the stock of knowledge.

It is also possible that the average productivity of research depends on the number of people searching for new ideas at any point in time. For example, perhaps duplication of effort is more likely when there are more persons engaged in research. One way of modeling this possibility is to suppose that it is really $L_A{}^{\lambda}$, where λ is some parameter between zero and one, rather than L_A that enters the production function for new ideas. This, together with equations (5.3) and (5.2), suggests focusing on the following general production function for ideas:

$$\dot{A} = \theta L_A^{\lambda} A^{\phi}. \tag{5.4}$$

For reasons that will become clear, we will assume that $\phi < 1$.

Equations (5.2) and (5.4) illustrate a very important aspect of modeling economic growth.[2] Individual researchers, being small relative to the economy as a whole, take $\bar{\theta}$ as given and see constant returns to research. As in equation (5.2), an individual engaged in research creates $\bar{\theta}$ new ideas. In the economy as a whole, however, the production function for ideas may not be characterized by constant returns to scale. While $\bar{\theta}$ will change by only a minuscule amount in response to the actions of a single researcher, it clearly varies with aggregate research effort.[3] For example,

[2]This modeling technique will be explored again in Chapter 9 in the context of "AK" models of growth.

[3]Notice that the exact expression for $\bar{\theta}$, incorporating both duplication and knowledge spillovers, is $\bar{\theta} = \theta L_A^{\lambda-1} A^{\phi}$.

$\lambda < 1$ may reflect an externality associated with duplication: some of the ideas created by an individual researcher may not be new to the economy as a whole. This is analogous to congestion on a highway. Each driver ignores the fact that his or her presence makes it slightly harder for other drivers to get where they are going. The effect of any single driver is negligible, but summed across all drivers, the effects can be important.

Similarly, the presence of A^{ϕ} is treated as external to the individual agent. Consider the case of $\phi > 0$, reflecting a positive knowledge spillover in research. The gains to society from the theory of gravitation far outweighed the benefit that Isaac Newton was able to capture. Much of the knowledge he created "spilled over" to future researchers. Of course, Newton himself also benefited from the knowledge created by previous scientists such as Kepler, as he recognized in the famous statement, "If I have seen farther than others, it is because I was standing on the shoulders of giants." With this in mind, we might refer to the externality associated with ϕ as the "standing on shoulders" effect, and by extension, the externality associated with λ as the "stepping on toes" effect.

Next, we need to discuss how resources are allocated in this economy. There are two key allocations. First, we assume (as before) that a constant fraction of output is invested in capital. Second, we have to decide how much labor works to produce output and how much works to produce ideas, recognizing that these two activities employ all of the labor in the economy:

$$L_Y + L_A = L.$$

In a more sophisticated model (and indeed, in Romer's original paper), the allocation of labor is determined by utility maximization and markets. However, it is again convenient to make the Solow-style assumption that the allocation of labor is constant; this assumption will be relaxed in Section 5.2. We assume that a constant fraction, $L_A/L = s_R$, of the labor force engages in R&D to produce new ideas, and the remaining fraction, $1 - s_R$, produces output.

Finally, the economy has some initial endowments when it begins. We assume the economy starts out with K_0 units of capital, L_0 units of labor, and A_0 ideas. This completes our setup of the model and we are ready to begin solving for some key endogenous variables, beginning with the long-run growth rate of this economy.

5.1 GROWTH IN THE ROMER MODEL

What is the growth rate in this model along a balanced growth path? Provided a constant fraction of the population is employed producing ideas (which we will show to be the case below), the model follows the neoclassical model in predicting that all per capita growth is due to technological progress. Letting lowercase letters denote per capita variables, and letting g_x denote the growth rate of some variable x along the balanced growth path, it is easy to show that

$$g_y = g_k = g_A.$$

That is, per capita output, the capital-labor ratio, and the stock of ideas must all grow at the same rate along a balanced growth path.[4] If there is no technological progress in the model, then there is no growth.

Therefore, the important question is "What is the rate of technological progress along a balanced growth path?" The answer to this question is found by rewriting the production function for ideas, equation (5.4). Dividing both sides of this equation by A yields

$$\frac{\dot{A}}{A} = \theta \frac{L_A^\lambda}{A^{1-\phi}}. \tag{5.5}$$

Along a balanced growth path, $\dot{A}/A \equiv g_A$ is constant. But this growth rate will be constant if and only if the numerator and the denominator of the right-hand side of equation (5.5) grow at the same rate. Taking logs and derivatives of both sides of this equation,

$$0 = \lambda \frac{\dot{L}_A}{L_A} - (1-\phi)\frac{\dot{A}}{A}. \tag{5.6}$$

Along a balanced growth path, the growth rate of the number of researchers must be equal to the growth rate of the population—if it were higher, the number of researchers would eventually exceed the

[4] To see this, follow the arguments we made in deriving equation (2.10) in Chapter 2. Intuitively, the capital-output ratio must be constant along a balanced growth path. Recognizing this fact, the production function implies that y and k must grow at the same rate as A.

population, which is impossible. That is, $\dot{L}_A/L_A = n$. Substituting this into equation (5.6) yields

$$g_A = \frac{\lambda n}{1 - \phi}. \tag{5.7}$$

Thus the long-run growth rate of this economy is determined by the parameters of the production function for ideas and the rate of growth of researchers, which is ultimately given by the population growth rate.

Several features of this equation deserve comment. First, what is the intuition for the equation? The intuition is most easily seen by considering the special case in which $\lambda = 1$ and $\phi = 0$ so that the productivity of researchers is the constant θ. In this case, there is no duplication problem in research and the productivity of a researcher today is independent of the stock of ideas that have been discovered in the past. The production function for ideas looks like

$$\dot{A} = \theta L_A.$$

Now suppose that the number of people engaged in the search for ideas is constant. Because θ is also constant, this economy generates a constant number of new ideas, θL_A, each period. To be more concrete, let's suppose $\theta L_A = 100$. The economy begins with some stock of ideas, A_0, generated by previous discoveries. Initially, the one hundred new ideas per period may be a large fraction of the existing stock, A_0. Over time, though, the stock grows, and the one hundred new ideas becomes a smaller and smaller fraction of the existing stock. Therefore, the *growth rate* of the stock of ideas falls over time, eventually approaching zero. Notice, however, that technological progress never ceases. The economy is always creating one hundred new ideas. It is simply that these one hundred new ideas shrink in comparison with the accumulated stock of ideas.

In order to generate exponential growth, the number of new ideas must be expanding over time. This occurs if the number of researchers is increasing—for example, because of world population growth. More researchers mean more ideas, sustaining growth in the model. In this case, the growth in ideas is clearly related to the growth in population, which explains the presence of population growth in equation (5.7).

It is interesting to compare this result to the effect of population growth in the neoclassical growth model. There, for example, a higher population growth rate reduces the level of income along a balanced growth path. More people means that more capital is needed to keep

K/L constant, but capital runs into diminishing returns. Here, an important additional effect exists. People are the key input to the creative process. A larger population generates more ideas, and because ideas are nonrivalrous, everyone in the economy benefits.

What evidence can be presented to support the contention that the per capita growth rate of the world economy depends on population growth? First, notice that this particular implication of the model is very difficult to test. We have already indicated that this model of the engine of growth is meant to describe the advanced countries of the world taken as a whole. Thus, we cannot use evidence on population growth *across* countries to test the model. In fact, we have already presented one of the most compelling pieces of evidence in Chapter 4. Recall the plot in Figure 4.4 of world population growth rates over the last two thousand years. Sustained and rapid population growth is a rather recent phenomenon, just as is sustained and rapid growth in per capita output. Increases in the rate of population growth from the very low rate observed over most of history occurred at roughly the same time as the Industrial Revolution.

The result that the growth rate of the economy is tied to the growth rate of the population implies another seemingly strong result: if the population (or at least the number of researchers) stops growing, long-run growth ceases. What do we make of this prediction? Rephrasing the question slightly, if research effort in the world were constant over time, would economic growth eventually grind to a halt? This model suggests that it would. A constant research effort cannot continue the proportional increases in the stock of ideas needed to generate long-run growth.

Actually, there is one special case in which a constant research effort can sustain long-run growth, and this brings us to our second main comment about the model. The production function for ideas considered in the original Romer (1990) paper assumes that $\lambda = 1$ and $\phi = 1$. That is,

$$\dot{A} = \theta L_A A.$$

Rewriting the equation slightly, we can see that this version of the Romer model *will* generate sustained growth in the presence of a constant research effort:

$$\frac{\dot{A}}{A} = \theta L_A. \tag{5.8}$$

In this case, Romer assumes that the productivity of research is proportional to the existing stock of ideas: $\bar{\theta} = \theta A$. With this assumption, the productivity of researchers grows over time, even if the number of researchers is constant.

The advantage of this specification, however, is also its drawback. World research effort has increased enormously over the last forty years and even over the last century (see Figure 4.6 in Chapter 4 for a reminder of this fact). Since L_A is growing rapidly over time, the original Romer formulation in equation (5.8) predicts that the growth rate of the advanced economies should also have risen rapidly over the last forty years or the last century. We know this is far from the truth. The average growth rate of the U.S. economy, for example, has been very close to 1.8 percent per year for the last hundred years. This easily rejected prediction of the original Romer formulation is avoided by requiring that ϕ is less than one, which returns us to the results associated with equation (5.7).[5]

Notice that nothing in this reasoning rules out increasing returns in research or positive knowledge spillovers. The knowledge spillover parameter, ϕ, may be positive and quite large. What the reasoning points out is that the somewhat arbitrary case of $\phi = 1$ is strongly rejected by empirical observation.[6]

Our last comment about the growth implications of this model of technology is that the results are similar to the neoclassical model in one important way. In the neoclassical model, changes in government policy and changes in the investment rate have no long-run effect on economic growth. This result was not surprising once we recognized that all growth in the neoclassical model was due to exogenous technological progress. In this model with endogenous technological progress, however, we have the same result. The long-run growth rate is invariant to changes in the investment rate, and even to changes in the share of the population that is employed in research. This is seen by noting that none of the parameters in equation (5.7) is affected when, say, the investment rate or the R&D share of labor is changed. Instead, these policies affect the growth rate along a transition path to the new steady

[5] This point is made in Jones (1995a).

[6] The same evidence also rules out values of $\phi > 1$. Such values would generate accelerating growth rates even with a constant population!

state altering the *level* of income. That is, even after we endogenize technology in this model, the long-run growth rate cannot be manipulated by policy makers using conventional policies such as subsidies to R&D.

5.2 GROWTH EFFECTS VERSUS LEVEL EFFECTS

The fact that standard policies cannot affect long-run growth is *not* a feature of the original Romer model, nor of many other idea-based growth models that followed, including the Schumpeterian growth models of Aghion and Howitt (1992) and Grossman and Helpman (1991). Much of the theoretical work in new growth theory has sought to develop models in which policy changes *can* have effects on long-run growth.

The idea-based models in which changes in policy can permanently increase the growth rate of the economy all rely on the assumption that $\phi = 1$, or its equivalent. As shown above, this assumption generates the counterfactual prediction that growth rates should accelerate over time with a growing population. Jones (1995a) generalized these models to the case of $\phi < 1$ to eliminate this defect and showed the somewhat surprising implication that this eliminates the long-run growth effects of policy as well. We will discuss these issues in more detail in Chapter 9.

5.3 COMPARATIVE STATICS: A PERMANENT INCREASE IN THE R&D SHARE

What happens to the advanced economies of the world if the share of the population searching for new ideas increases permanently? For example, suppose there is a government subsidy for R&D that increases the fraction of the labor force doing research.

An important feature of the model we have just developed is that many policy changes (or comparative statics) can be analyzed with techniques we have already developed. Why? Notice that technological progress in the model can be analyzed by itself—it doesn't depend on capital or output, but only on the labor force and the share of the population devoted to research. Once the growth rate of A is constant, the model behaves just like the Solow model with exogenous technological progress. Therefore, our analysis proceeds in two steps. First, we consider what happens to technological progress and to the stock of ideas after the increase in R&D intensity occurs. Second, we analyze the model as we did the Solow model, in steps familiar from Chapter 2. Before we proceed, it is worth noting that the analysis of changes that do not affect technology, such as an increase in the investment rate, is exactly like the analysis of the Solow model.

Now consider what happens if the share of the population engaged in research increases permanently. To simplify things slightly, let's assume that $\lambda = 1$ and $\phi = 0$ again; none of the results are qualitatively affected by this assumption. It is helpful to rewrite equation (5.5) as

$$\frac{\dot{A}}{A} = \theta \frac{s_R L}{A}, \tag{5.9}$$

where s_R is the share of the population engaged in R&D, so that $L_A = s_R L$.

Figure 5.1 shows what happens to technological progress when s_R increases permanently to s'_R, assuming the economy begins in steady state. In steady state, the economy grows along a balanced growth path at the rate of technological progress, g_A, which happens to equal the rate of population growth under our simplifying assumptions. The ratio L_A/A is therefore equal to g_A/θ. Suppose the increase in s_R occurs at time $t = 0$. With a population of L_0, the number of researchers increases as s_R increases, so that the ratio L_A/A jumps to a higher level. The additional researchers produce an increased number of new ideas, so the growth rate of technology is also higher at this point. This situation corresponds to the point labeled "X" in the figure. At X, technological progress $\dot{A}/A$ exceeds population growth n, so the ratio L_A/A declines over time, as indicated by the arrows. As this ratio declines, the rate of technological change gradually falls also, until the economy returns to the balanced growth path where $g_A = n$. Therefore, a permanent increase in the share of the population devoted to research raises the rate of technological progress temporarily, but not in the long run. This behavior is depicted in Figure 5.2.

What happens to the level of technology in this economy? Figure 5.3 answers this question. The level of technology is growing along a balanced growth path at rate g_A until time $t = 0$. At this time, the growth rate increases and the level of technology rises faster than before. Over time, however, the growth rate falls until it returns to g_A.

FIGURE 5.1 **TECHNOLOGICAL PROGRESS: AN INCREASE IN THE R&D SHARE**

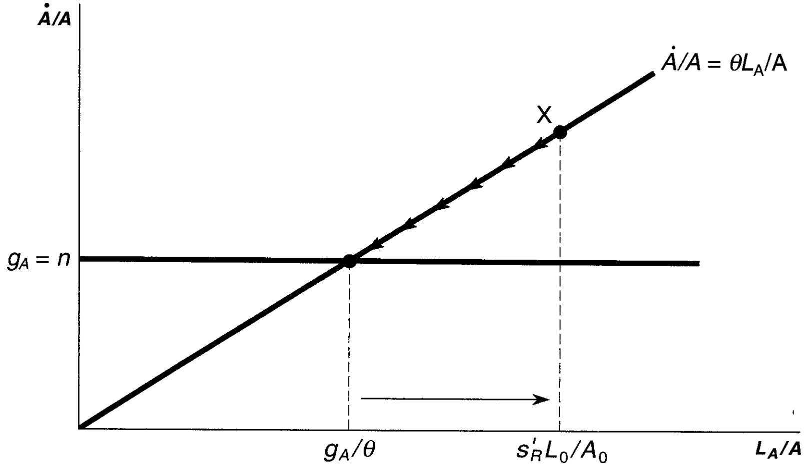
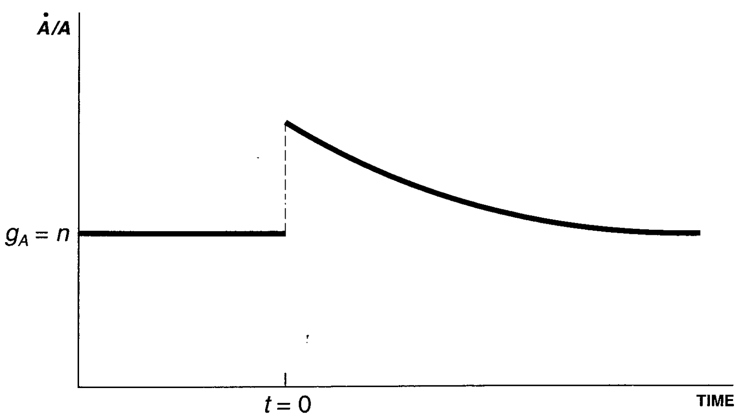
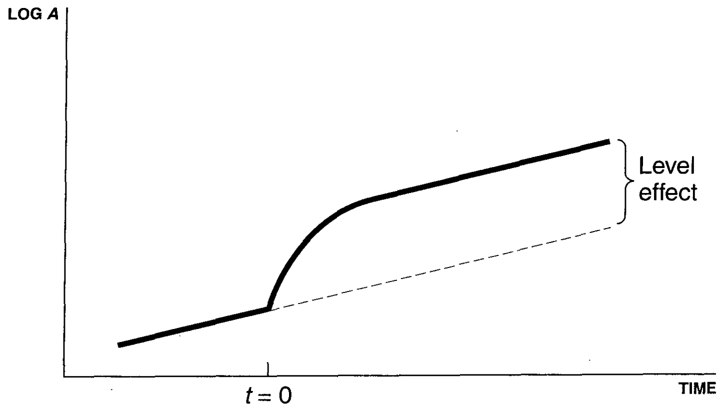


FIGURE 5.2 **$\dot{A}/A$ OVER TIME**



The *level* of technology is permanently higher as a result of the permanent increase in R&D. Notice that a permanent increase in s_R in the Romer model generates transition dynamics that are qualitatively similar to the dynamics generated by an increase in the investment rate in the Solow model.

FIGURE 5.3 **THE LEVEL OF TECHNOLOGY OVER TIME**



Now that we know what happens to technology over time, we can analyze the remainder of the model in a Solow framework. The long-run growth rate of the model is constant, so much of the algebra that we used in analyzing the Solow model applies. For example, the ratio y/A is constant along a balanced growth path and is given by an equation similar to equation (2.13):

$$\left(\frac{y}{A}\right)^* = \left(\frac{s_K}{n + g_A + \delta}\right)^{\alpha/(1-\alpha)} (1 - s_R). \tag{5.10}$$

The only difference is the presence of the term $1 - s_R$, which adjusts for the difference between output per worker, L_Y, and output per capita, L.

Notice that along a balanced growth path, equation (5.9) can be solved for the level of A in terms of the labor force:

$$A = \frac{\theta s_R L}{g_A}.$$

Combining this equation with (5.10), we get

$$y^*(t) = \left(\frac{s_K}{n + g_A + \delta}\right)^{\alpha/(1-\alpha)} (1 - s_R) \frac{\theta s_R}{g_A} L(t). \tag{5.11}$$

In this simple version of the model, per capita output is proportional to the population of the (world) economy along a balanced growth path. In other words, the model exhibits a *scale effect* in levels: a larger world economy will be a richer world economy. This scale effect arises fundamentally from the nonrivalrous nature of ideas: a larger economy provides a larger market for an idea, raising the return to research (a demand effect). In addition, a more populous world economy simply has more potential creators of ideas in the first place (a supply effect).

The other terms in equation (5.11) are readily interpreted. The first term is familiar from the original Solow model. Economies that invest more in capital will be richer, for example. Two terms involve the share of labor devoted to research, s_R. The first time s_R appears, it enters negatively to reflect the fact that more researchers mean fewer workers producing output. The second time, it enters positively to reflect the fact that more researchers mean more ideas, which increases the productivity of the economy.

5.2 THE ECONOMICS OF THE ROMER MODEL

The first half of this chapter has analyzed the Romer model without discussing the economics underlying the model. A number of economists in the 1960s developed models with similar macroeconomic features.[7] However, the development of the microfoundations of such models had to wait until the 1980s when economists better understood how to model imperfect competition in a general equilibrium setting.[8] In fact, one of the important contributions of Romer (1990) was to explain exactly how to construct an economy of profit-maximizing agents that endogenizes technological progress. The intuition behind this insight was developed in Chapter 4. Developing the mathematics is the subject of the remainder of this section. Because this section is somewhat difficult, some readers may wish to skip to Section 5.3.

[7]For example, Uzawa (1965), Phelps (1966), Shell (1967), and Nordhaus (1969).

[8]Key steps in this understanding were accomplished by Spence (1976), Dixit and Stiglitz (1977), and Ethier (1982).

The Romer economy consists of three sectors: a final-goods sector, an intermediate-goods sector, and a research sector. The reason for two of the sectors should be clear: some firms must produce output and some firms must produce ideas. The reason for the intermediate-goods sector is related to the presence of increasing returns discussed in Chapter 4. Each of these sectors will be discussed in turn. Briefly, the research sector creates new ideas, which take the form of new varieties of capital goods—new computer chips, industrial robots, or printing presses. The research sector sells the exclusive right to produce a specific capital good to an intermediate-goods firm. The intermediate-goods firm, as a monopolist, manufactures the capital good and sells it to the final-goods sector, which produces output.

5.2.1 THE FINAL-GOODS SECTOR

The final-goods sector of the Romer economy is very much like the final-goods sector of the Solow model. It consists of a large number of perfectly competitive firms that combine labor and capital to produce a homogeneous output good, Y. The production function is specified in a slightly different way, though, to reflect the fact that there is more than one capital good in the model:

$$Y = L_Y^{1-\alpha} \sum_{j=1}^{A} x_j^{\alpha}.$$

Output, Y, is produced using labor, L_Y, and a number of different capital goods, X_j, which we will also call "intermediate goods." At any point in time, A measures the number of capital goods that are available to be used in the final-goods sector, and firms in the final-goods sector take this number as given. Inventions or ideas in the model correspond to the creation of new capital goods that can be used by the final-goods sector to produce output.

Notice that this production function can be rewritten as

$$Y = L_Y^{1-\alpha} x_1^{\alpha} + L_Y^{1-\alpha} x_2^{\alpha} + \cdots + L_Y^{1-\alpha} x_A^{\alpha},$$

and it is easy to see that, for a given A, the production function exhibits constant returns to scale; doubling the amount of labor and the amount of each capital good will exactly double output.

It turns out for technical reasons to be easier to analyze the model if we replace the summation in the production function with an integral:

$$Y = L_Y^{1-\alpha} \int_0^A x_j^{\alpha} dj.$$

Then, A measures the range of capital goods that are available to the final-goods sector, and this range is the interval on the real line $[0, A]$. The basic interpretation of this equation, though, is unaffected by this technicality.

With constant returns to scale, the number of firms cannot be pinned down, so we will assume there are a large number of identical firms producing final output and that perfect competition prevails in this sector. We will also normalize the price of the final output, Y, to unity.

Firms in the final-goods sector have to decide how much labor and how much of each capital good to use in producing output. They do this by solving the profit-maximization problem:

$$\max_{L_Y, x_j} L_Y^{1-\alpha} \int_0^A x_j^{\alpha} dj - wL_Y - \int_0^A p_j x_j dj,$$

where p_j is the rental price for capital good j and w is the wage paid for labor. The first-order conditions characterizing the solution to this problem are

$$w = (1 - \alpha)\frac{Y}{L_Y} \tag{5.12}$$

and

$$p_j = \alpha L_Y^{1-\alpha} x_j^{\alpha-1}, \tag{5.13}$$

where this second condition applies to each capital good j. The first condition says that firms hire labor until the marginal product of labor equals the wage. The second condition says the same thing, but for capital goods: firms rent capital goods until the marginal product of each kind of capital equals the rental price, p_j. To see the intuition for these equations, suppose the marginal product of a capital good were higher than its rental price. Then the firm should rent another unit; the output produced will more than pay for the rental price. If the marginal product is below the rental price, then the firm can increase profits by reducing the amount of capital used.

5.2.2 THE INTERMEDIATE-GOODS SECTOR

The intermediate-goods sector consists of monopolists who produce the capital goods that are sold to the final-goods sector. These firms gain their monopoly power by purchasing the design for a specific capital good from the research sector. Because of patent protection, only one firm manufactures each capital good.

Once the design for a particular capital good has been purchased (a fixed cost), the intermediate-goods firm produces the capital good with a very simple production function: one unit of raw capital can be automatically translated into one unit of the capital good. The profit maximization problem for an intermediate goods firm is then

$$\max_{x_j} \pi_j = p_j(x_j)x_j - rx_j,$$

where $p_j(x)$ is the demand function for the capital good given in equation (5.13). The first-order condition for this problem, dropping the j subscripts, is

$$p'(x)x + p(x) - r = 0.$$

Rewriting this equation we get

$$p'(x)\frac{x}{p} + 1 = \frac{r}{p},$$

which implies that

$$p = \frac{1}{1 + \frac{p'(x)x}{p}} r.$$

Finally, the elasticity, $p'(x)x/p$, can be calculated from the demand curve in equation (5.13). It is equal to $\alpha - 1$, so the intermediate-goods firm charges a price that is simply a markup over marginal cost, r:

$$p = \frac{1}{\alpha} r.$$

This is the solution for each monopolist, so that all capital goods sell for the same price. Because the demand functions in equation (5.13) are also the same, each capital good is employed by the final-goods firms in the same amount: $x_j = x$. Therefore, each capital-goods firm earns the same profit. With some algebra, one can show that this profit is given by

$$\pi = \alpha(1 - \alpha)\frac{Y}{A}. \tag{5.14}$$

Finally, the total demand for capital from the intermediate-goods firms must equal the total capital stock in the economy:

$$\int_0^A x_j dj = K.$$

Since the capital goods are each used in the same amount, x, this equation can be used to determine x:

$$x = \frac{K}{A}. \tag{5.15}$$

The final-goods production function can be rewritten, using the fact that $x_j = x$, as

$$Y = AL_Y^{1-\alpha}x^\alpha,$$

and substituting from equation (5.15) reveals that

$$\begin{aligned} Y &= AL_Y^{1-\alpha}A^{-\alpha}K^\alpha \\ &= K^\alpha(AL_Y)^{1-\alpha}. \end{aligned} \tag{5.16}$$

That is, we see that the production technology for the final-goods sector generates the same aggregate production function used throughout this book. In particular, this is the aggregate production function used in equation (5.1).

THE RESEARCH SECTOR

Much of the analysis of the research sector has already been provided. The research sector is essentially like gold mining in the wild West in the mid-nineteenth century. Anyone is free to "prospect" for ideas, and the reward for prospecting is the discovery of a "nugget" that can be sold. Ideas in this model are designs for new capital goods: a faster computer chip, a method for genetically altering corn to make it more resistant to pests, or a new way to organize movie theaters. These designs can be thought of as instructions that explain how to transform a unit of raw capital into a unit of a new capital good. New designs are discovered according to equation (5.4).

When a new design is discovered, the inventor receives a patent from the government for the exclusive right to produce the new capital good. (To simplify the analysis, we assume that the patent lasts forever.) The inventor sells the patent to an intermediate-goods firm and uses the proceeds to consume and save, just like any other agent in the model. But what is the price of a patent for a new design?

We assume that anyone can bid for the patent. How much will a potential bidder be willing to pay? The answer is the present discounted value of the profits to be earned by an intermediate-goods firm. Any less, and someone would be willing to bid higher; any more, and no one would be willing to bid. Let P_A be the price of a new design, this present discounted value. How does P_A change over time? The answer lies in an extremely useful line of reasoning in economics and finance called the method of *arbitrage*.

The arbitrage argument goes as follows. Suppose we have some money to invest for one period. We have two options. First, we can put the money in the "bank" (in this model, this is equivalent to purchasing a unit of capital) and earn the interest rate r. Alternatively, we can purchase a patent for one period, earn the profits that period, and then sell the patent. In equilibrium, it must be the case that the rate of return from both of these investments is the same. If not, everyone would jump at the more profitable investment, driving its return down. Mathematically, the *arbitrage equation* states that the returns are the same:

$$rP_A = \pi + \dot{P}_A. \tag{5.17}$$

The left-hand side of this equation is the interest earned from investing P_A in the bank; the right-hand side is the profits plus the capital gain or loss that results from the change in the price of the patent. These two must be equal in equilibrium.

Rewriting equation (5.17) slightly,

$$r = \frac{\pi}{P_A} + \frac{\dot{P}_A}{P_A}.$$

Along a balanced growth path, r is constant.[9] Therefore, π/P_A must be constant also, which means that π and P_A have to grow at the same rate; this rate turns out to be the population growth rate, n.[10] Therefore, the arbitrage equation implies that

$$P_A = \frac{\pi}{r - n}. \tag{5.18}$$

This equation gives the price of a patent along a balanced growth path.

5.2.4 SOLVING THE MODEL

We have now described the market structure and the microeconomics underlying the basic equations given in Section 5.1. The model is somewhat complicated, but several features that were discussed in Chapter 4 are worth noting. First, the aggregate production function exhibits increasing returns. There are constant returns to K and L, but increasing returns once we note that ideas, A, are also an input to production. Second, the increasing returns require imperfect competition. This appears in the model in the intermediate-goods sector. Firms in this sector are monopolists, and capital goods sell at a price that is greater than marginal cost. However, the profits earned by these firms are extracted by the inventors, and these profits simply compensate the inventors for the time they spend "prospecting" for new designs. This framework is called *monopolistic competition*. There are no economic profits in the model; all rents compensate some factor input. Finally, once we depart from the world of perfect competition there is no reason to think that markets yield the "best of all possible worlds." This last point is one that we develop more carefully in the final section of this chapter.

We have already solved for the growth rate of the economy in steady state. The part of the model that remains to be solved is the allocation of labor between research and the final-goods sector. Rather than assuming s_R is constant, we let it be determined endogenously by the model.

[9] The interest rate r is constant for the usual reasons. It will be the price at which the supply of capital is equal to the demand for capital, and it will be proportional to Y/K.
[10] To see this, recall from equation (5.14) that π is proportional to Y/A. Per capita output, y, and A grow at the same rate, so that Y/A will grow at the rate of population growth.

Once again, the concept of arbitrage enters. It must be the case that, at the margin, individuals in this simplified model are indifferent between working in the final-goods sector and working in the research sector. Labor working in the final-goods sector earns a wage that is equal to its marginal product in that sector, as given in equation (5.12):

$$w_Y = (1 - \alpha)\frac{Y}{L_Y}.$$

Researchers earn a wage based on the value of the designs they discover. We will assume that researchers take their productivity in the research sector, $\bar{\theta}$, as given. They do not recognize that productivity falls as more labor enters because of duplication, and they do not internalize the knowledge spillover associated with ϕ. Therefore, the wage earned by labor in the research sector is equal to its marginal product, $\bar{\theta}$, multiplied by the value of the new ideas created, P_A:

$$w_R = \bar{\theta} P_A.$$

Because there is free entry into both the research sector and the final-goods sector, these wages must be the same: $w_Y = w_R$. This condition, with some algebra shown in the appendix to this chapter, reveals that the share of the population that works in the research sector, s_R, is given by

$$s_R = \frac{1}{1 + \frac{r-n}{\alpha g_A}}. \tag{5.19}$$

Notice that the faster the economy grows (the higher is g_A), the higher the fraction of the population that works in research. The higher the discount rate that applies to current profits to get the present discounted value $(r - n)$, the lower the fraction working in research.[11]

With some algebra, one can show that the interest rate in this economy is given by $r = \alpha^2 Y/K$. Notice that this is *less* than the marginal product of capital, which from equation (5.16) is the familiar $\alpha Y/K$. This difference reflects an important point. In the Solow model with perfect competition and constant returns to scale, all factors are paid their marginal products: $r = \alpha Y/K$, $w = (1 - \alpha)Y/L$, and therefore $rK + wL = Y$. In the Romer model, however, production in the

[11] One can eliminate the interest rate from this equation by noting that $r = \alpha^2 Y/K$ and getting the capital-output ratio from the capital accumulation equation: $Y/K = (n + g + \delta)/s_K$.

economy is characterized by increasing returns and all factors cannot be paid their marginal products. This is clear from the Solow example just given: because $rK + wL = Y$, there is no output in the Solow economy remaining to compensate individuals for their effort in creating new A. *This* is what necessitates imperfect competition in the model. Here, capital is paid less than its marginal product, and the remainder is used to compensate researchers for the creation of new ideas.

That completes the equilibrium solution for the Romer model. The key point was to have the market allocate resources, with the key allocation being the decision regarding how much labor to use in research versus production. As we showed, it was the profits associated with intermediate good firms that gave value to patents for new varieties, and in turn made research worth doing. The profits thus provide the return to research that is crucial to sustained economic growth.

GROWTH THROUGH CREATIVE DESTRUCTION

We set out in this chapter to develop an explicit theory of technological progress. The Romer model viewed technological progress as an increase in the number of intermediate goods, and showed how this increase could come about as the result of profit-maximizing behavior by innovators and firms.

One thing to note about the Romer model is that, once invented, each variety of intermediate good continues to be used forever. If we applied this strictly then we would expect to see steam engines, for instance, used alongside electric motors. An alternative type of endogenous growth theory explicitly allows for an innovation to *replace* an existing intermediate good in the production process.

Models that feature such quality improvements in intermediate goods were developed by Aghion and Howitt (1992) and Grossman and Helpman (1991). The former coined the term "Schumpeterian" to describe their model. Joseph Schumpeter, writing in the late 1930s and early 1940s, discussed capitalism as a process of *creative destruction*, in which existing businesses and technologies are replaced by new ones. Growth required the continual obsolescence of old techniques as new ones were invented, improving the productivity of the economy at each step.

The model we develop in this section will attempt to capture those elements, and you will see that while many of the long-run results will be similar to the Romer model, this type of model has other unique results that arise when today's innovators realize that they too will someday be replaced.

5.3.1 THE BASIC ELEMENTS OF THE SCHUMPETERIAN MODEL

Similar to the approach with the Romer model, we'll begin by looking at the overall structure of the Schumpeterian model before turning to the market structure that lay behind it. The process of innovation is similar to that used in Segerstrom (1998), which will keep the model consistent with the empirical facts discussed in Section 5.1.[12]

The aggregate production function for the Schumpeterian model looks similar to the Solow or Romer function,

$$Y = K^{\alpha}(A_i L_Y)^{1-\alpha}, \tag{5.20}$$

with one particular difference. Note that what we've called the stock of ideas, A, is indexed by i. This i indexes ideas, and as i gets larger, A_i gets larger.

You can think of the A_i term as capturing the latest available technology. A_4 could represent modern cars, while A_3 is the Model T Ford, A_2 is a horse cart, and A_1 is walking. Each time we innovate, we get more productive, as in the Romer model. However, innovation here is occurring in steps, rather than continuously.

Because innovation occurs in steps, we cannot write down an equation exactly like (5.4), and we have to break the growth in A down into two parts: the size of innovations when they occur, and the chance that an innovation happens.

In the Schumpeterian model, the size of innovations is held constant, although that is not crucial to the results we will develop. Let

$$A_{i+1} = (1 + \gamma)A_i, \tag{5.21}$$

where γ captures the "step size," or the amount that productivity rises when an innovation actually occurs.

Growth in this economy occurs only when an innovation happens, and these don't always happen. The growth rate of A, *from innovation to innovation,* is

$$\frac{A_{i+1} - A_i}{A_i} = \gamma. \tag{5.22}$$

[12]Segerstrom, Anant, and Dinopoulos (1990) and Kortum (1997) also provide models with similar properties.

Note that this is not the growth rate of A_i over *time*. That depends on how often these changes in A occur in time, and to know that, we need to know about the chances that innovation happens.

The chance of an innovation will depend on research effort. For any individual doing research, let his or her probability of discovering the next innovation be $\bar{\mu}$ at any moment in time. This term is taken by the individual as given, but will be subject to similar forces that affect innovation in the Romer model. Here, though, the "standing on shoulders" and "stepping on toes" will affect the probability of innovation, not the size of innovation. To be more specific, let

$$\bar{\mu} = \theta \frac{L_A^{\lambda - 1}}{A_i^{1-\phi}}. \tag{5.23}$$

For the economy as a whole, the probability of an innovation occurring at any moment in time is equal to the individual probability of innovation, $\bar{\mu}$, times the number of individuals doing research:

$$P(innovation) = \bar{\mu} L_A = \theta \frac{L_A^{\lambda} A_i^{\phi}}{A_i}. \tag{5.24}$$

This probability involves two effects of A_i. With $0 < \phi < 1$, increasing A_i increases the chance of finding a new innovation, the typical standing on shoulders effect. However, the probability of making new innovations is lower as A_i gets larger, as in Segerstrom (1998). To push the analogy, standing on shoulders allows researchers to see more possible opportunities, but it also means they are seeing possibilities increasingly far away.

Aside from the process of technological change, the remaining parts of the Schumpeterian model are identical to the Romer model. Specifically, capital accumulates through

$$\dot{K} = s_K Y - \delta K,$$

while the total labor force rises exponentially,

$$\frac{\dot{L}}{L} = n,$$

and that labor force is divided between workers in the final goods sector (L_Y) and researchers (L_A):

$$L = L_Y + L_A.$$

Initially, the economy has some capital stock K_0 and total labor force L_0. We will assume that the initial technology level is A_0, meaning only that the initial technology level is indexed to zero, and growth will consist of advancing to the next "step" in the ladder of technology.

5.3.2 GROWTH IN THE SCHUMPETERIAN MODEL

The growth rate of this economy, with respect to time, is not immediately obvious. As innovations only occur randomly, there will be periods of time in which output per capita is not growing at all, followed by distinct jumps when innovations occur. Because of the random arrival of innovations, we cannot specify the precise path that income per capita will take.

However, we can say something about growth over long periods of time. We have a standard neoclassical model, given our production function, and our standard assumptions regarding capital accumulation and population growth. Given these, we can conceive of a balanced growth path where the *average* growth rates of output per capita (g_y) and the capital-labor ratio (g_k) are constant and equal to the *average* growth ate of productivity (g_A).

At any given moment, we have a probability of innovating, $\bar{\mu}L_A$, and we know the size of the innovation that will occur if successful, γ. The expected growth rate of A over *time* is

$$E\left[\frac{\dot{A}}{A}\right] = \gamma\bar{\mu}L_A = \gamma\theta\frac{L_A^\lambda}{A_i^{1-\phi}}. \qquad (5.25)$$

If we look over very long periods of time, then by the law of large numbers, the actual average growth rate will approach this expectation, so that

$$g_y = g_k = g_A = E\left[\frac{\dot{A}}{A}\right].$$

As in the Solow or Romer models, the trend growth rate of output per capita is governed by the growth rate of technology. Here, it so happens to be the expected value of the growth rate of technology.[13]

[13]The equivalence of growth rates to the expected value of growth in A is only approximate. If we allowed for a continuum of sectors, each experiencing Schumpeterian technological change, then the random arrival of innovations across sectors would even out across the sectors, and the equivalence of growth rates to $E[\dot{A}/A]$ would be exact.

Using this, we can again ask the (slightly modified) question, "What is the *expected* rate of technological progress along a balanced growth path?" The analysis now follows Section 5.1.1 very closely. To find the growth rate, taking logs and time derivatives of both sides of equation (5.25), we have that

$$0 = \lambda \frac{\dot{L}_A}{L_A} - (1 - \phi) E\left[\frac{\dot{A}_i}{A_i}\right], \tag{5.26}$$

where we've replaced the growth rate of A_i with its expectation.

As before, note that $\dot{L}_A / L_A = n$, otherwise the number of researchers would become larger than the population. Using this, we can solve equation (5.26) for the average growth rate

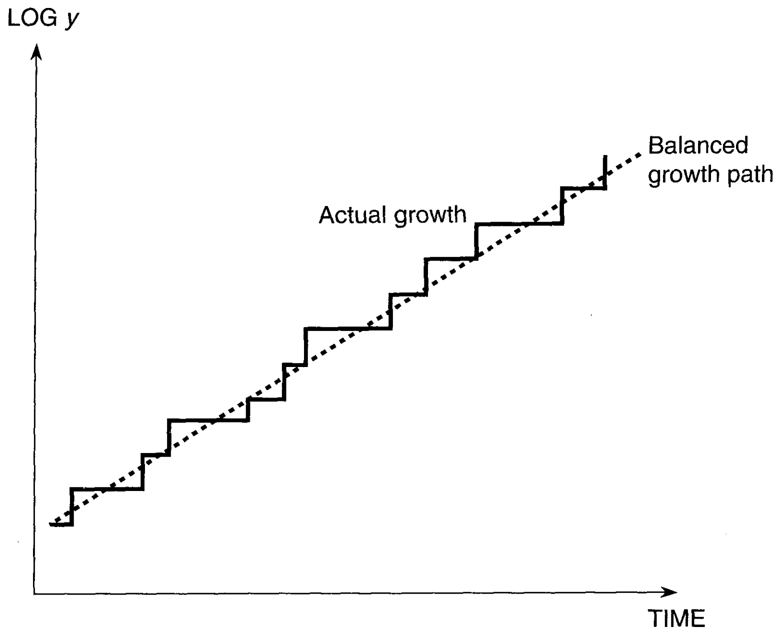
$$g_A = \frac{\lambda n}{1 - \phi}. \tag{5.27}$$

The average long-run growth rate in the Schumpeterian model is identical to that of the Romer model. As noted before, the *actual* growth rate of the economy won't be precisely this rate for any small period of time, because innovations arrive randomly. However, on average, the economy will grow at a rate dictated by the growth rate of population as well as the parameters governing the duplication of research effort (λ) and spillovers (ϕ).

Figure 5.4 shows the distinction between the average growth rate along the balanced growth path and the actual growth of income per capita. The bold line shows how log income per capita actually changes over time. There are flat sections, implying that no innovations have been made. When someone does discover the next innovation, log income per capita jumps upward by the amount γ. On average, income per capita is growing along the line labeled "Balanced growth path," which given equation (5.27), depends on the population growth rate.

It is interesting to note that γ, the size of each individual innovation, does not feature in the growth rate along the balanced growth path, and it is worth asking why. A larger γ introduces a larger boost to technology each time an innovation occurs, and if the probability of an innovation remains the same, then this should raise the growth rate. However, larger "steps" in innovation also raise the absolute size of A, which slows down the probability of finding the next innovation given our assumption that $\phi < 1$. As each innovation occurs, that next big breakthrough takes longer, which offsets the positive effect of a larger γ.

FIGURE 5.4 INCOME PER CAPITA ALONG BALANCED GROWTH PATH, SCHUMPETERIAN MODEL



5.4 THE ECONOMICS OF SCHUMPETERIAN GROWTH

As we did with the Romer model, we can explore the economics underlying the Schumpeterian model. This will again involve a model of imperfect competition, which as discussed earlier, is necessary in order to generate profits that can be used to compensate the researchers for their work. The differences in the models lie in how intermediate goods are used, and in the nature of innovation. These will lead ultimately to the Schumpeterian model having a different solution for the proportion of people engaged in research, and this will have some interesting implications for the role of competition in the economy.

We again have a final-goods sector, an intermediate-goods sector, and a research sector. Here, though, there will only be a single intermediate good, produced by a single monopolistic intermediate-goods firm that owns the patent. They produce the capital good that is used by the final-goods sector to produce output. The research sector consists of individuals who are trying to generate a new version of the capital good, one that is more productive for the final-goods sector firms to use. The research sector can sell the patent for their design to a new intermediate-goods firm that will then monopolize the market for the intermediate goods until they are replaced.

In this way, the model embodies the idea of "creative destruction" that came from Schumpeter originally. The intermediate-goods supplier is always in danger of being replaced by a new supplier, and this will play into the value that the intermediate-goods firm will pay for a patent.

5.4.1 THE FINAL-GOODS SECTOR

Whereas in the Romer model there were A intermediate goods used in production, here there is only one. The production function for final goods is specified as

$$Y = L_Y^{1-\alpha} A_i^{1-\alpha} x_i^{\alpha}.$$

Here, output Y is produced using labor, L_Y, as well as a single capital good, x_i, which as before can also be referred to as an intermediate good. Again, there are a large number of perfectly competitive firms in the final-goods sector. This production function remains constant returns to scale, as doubling the amount of capital goods and labor will produce exactly double the output.

Most crucially, note that the capital good, x_i, as well as the productivity term, A_i, are indexed by i. The i refers to the version of the capital good in use, and each version of the capital good comes with its own productivity level. If the final-goods firms use capital good x_i, then they are implicitly using the level of productivity A_i. Intuitively, one can think of x_i as representing how many units of a machine are being used, and A_i as representing how efficient those machines are. For example, $i = 1$ may be an old IBM mainframe computer, with a productivity of A_1. A modern server is $i = 2$, and has a productivity of $A_2 > A_1$. Even if the firms use the same number of each, so that $x_1 = x_2$, they will produce more output by using the servers as opposed to mainframes.

Innovations raise output only if final-goods firms actually purchase the latest version of the capital good. As we'll see below, intermediate-goods firms will sell all the versions of the capital goods at the same price. Therefore, final-good firms will purchase only the latest version, as it gives them the highest productivity level. In this way, the economy will always be operating with the latest technology. It's possible to have a more complex Schumpeterian model that includes the possibility of different versions of capital goods being used at the same time, but all the insights we develop here will still follow.

We can examine the demand of the final-goods firms for the intermediate good. They solve the profit maximization problem

$$\max_{L_Y,\, x_i} L_Y^{1-\alpha} A_i^{1-\alpha} x_i^{\alpha} - wL_Y - p_i x_i,$$

where ω is the wage for a unit of labor, and p_i is the rental price for a unit of x_i. The first-order conditions are standard, and show

$$w = (1 - \alpha)\frac{Y}{L_Y} \tag{5.28}$$

and

$$p_i = \alpha L_Y^{1-\alpha} A_i^{1-\alpha} x_i^{\alpha-1} \tag{5.29}$$

The first shows that firms hire labor until its marginal product is equal to the wage, and the second says that capital goods are purchased until their marginal product is equal to the price charged by the intermediate-goods firm.

5.4.2 THE INTERMEDIATE-GOODS SECTOR

An intermediate-goods firm is a monopolist that produces a single version of the capital good. They are monopolists because they have bought a design from the research sector, and patent protections ensure that no one else can produce their version.

As in the Romer model, an intermediate-goods firm will produce the capital good in a very simple manner: one unit of raw capital can be transformed into one unit of the capital good. The profit-maximization problem for the intermediate-goods firm is

$$\max_{x_i} \pi_i = p_i(x_i)x_i - rx_i,$$

where we note that the maximization problem is indexed by i, the version of the capital-good design that the firm owns. $p_i(x_i)$ is the demand function for the capital good from the final-goods sector (5.29).

The first-order condition for any firm i is.

$$p_i'(x_i)x_i + p_i(x_i) - r = 0,$$

and we've kept the i subscript in explicitly because we will want to highlight that the price charged by each firm will be identical. Similar to Section 5.2.2, we can rewrite this first-order condition as

$$p_i = \frac{1}{1 + \frac{p_i'(x_i)x_i}{p_i}} r.$$

The elasticity of demand for capital good x_i in the denominator can be found from equation (5.29). It is equal to $\alpha - 1$, so that any intermediate-goods firm charges

$$p_i = \frac{1}{\alpha} r,$$

a constant markup over the cost of producing the intermediate good.

This provides us with some insight into why final-goods firms only ever purchase one version of the capital good, and why that is the latest version. Since each intermediate-goods firm charges the same for a unit of the capital good, buying an old version of the capital good is as expensive as buying the latest version. Because the productivity is highest with the latest version, final-goods firms will always want to buy it over any others. This means that the economy is always operating with version i, and never with version $i - 1$ or $i - 2$ of the capital good.[14]

[14]This result holds strictly provided that the innovations are "drastic," meaning that γ is large enough that even if the old monopolist only charged marginal cost, r, for each unit of the old capital good, final-goods producers would still buy the new capital good. If innovations are "nondrastic," meaning that γ is relatively small, then the new monopolist can still drive the old monopolist out of business but will have to lower the price of new capital goods to less than r/α. This lower markup over marginal cost makes being a monopolist less profitable, which will reduce the incentive to innovate. This will have a level effect on output by lowering the fraction s_R, but it won't affect the growth rate along the balanced growth path.

Given that final-goods firms only buy the latest version of the capital good, only one intermediate-goods firm, the one that owns the patent to version i, will operate. The firm's profits are given by

$$\pi = \alpha(1 - \alpha)Y, \tag{5.30}$$

which is similar to the profits for a firm in the Romer model. However, here profits are not divided over multiple intermediate-goods firms, and so this is not divided by A, as it was in equation (5.14)

Finally, given only one intermediate-goods firm, it must be that all the capital in the economy is used to produce the latest version of the intermediate good, so that $x_i = K$. This means that aggregate output is

$$Y = K^{\alpha}(A_i L_Y)^{1-\alpha},$$

which is the same aggregate production function used throughout the book. The one distinction is that aggregate productivity is A_i, and not simply A. That is, productivity depends upon exactly which version of the capital good we are using. As discussed, A does not rise smoothly over time, but jumps when someone innovates and we move from capital good i to capital good $i + 1$. This occurs through the research sector described next.

5.4.3 THE RESEARCH SECTOR

The main distinction between the Schumpeterian model and the Romer model comes in how we conceive of innovation. In the Romer model, people prospected for new intermediate goods, and these arrived at a constant rate, given by equation (5.4). Here, everyone who does research is working on the same idea—version $i + 1$ of the capital good. An individual who is doing research has a constant probability of discovering this new version, denoted by $\bar{\mu}$.

If an inventor does discover a new version, he or she receives a patent from the government, and again we presume that this patent lasts forever. The inventor will again sell the patent to an intermediate-goods firm. This will be a new intermediate-goods firm. The existing intermediate-goods firm that produces version i will not purchase the patent for the version $i + 1$. We'll discuss below why this is true.

We will again use the idea of arbitrage to describe the value of the patent, P_A, to the intermediate-goods firm,

$$rP_A = \pi + \dot{P}_A - \bar{\mu}\ L_A P_A. \tag{5.31}$$

What differs from the Romer model is that the patent for a design in the Schumpeterian model will eventually lose all of its value. Recall that only the latest version of the capital good is ever used in production. If you own the patent for version i, then once someone invents version $i + 1$, you will be out of business. This is captured by the final term in the arbitrage equation. This says that with L_A people doing research, each with a probability $\bar{\mu}$ of innovating, then there is an $\bar{\mu}L_A$ chance of being replaced as the capital-goods provider. If you are replaced, then you lose the entire value of my patent, P_A.

Rearranging the arbitrage equation, we have

$$r = \frac{\pi}{P_A} + \frac{\dot{P}_A}{P_A} - \bar{\mu}\ L_A.$$

Along a balanced growth path, it must be that r is constant. The value of $\bar{\mu}L_A$ is the probability of a new innovation occurring, and this is constant along the balanced growth path as well. Let $\mu = \bar{\mu}\ L_A$ denote this aggregate probability.

The ratio π / P_A is therefore constant along the balanced growth path as well, so π and P_A must grow at the same rate. Given equation (5.30), we know that profits are proportional to aggregate output, which grows at the rate $g_y + n$. From our prior analysis of the model, we know that $g_y = \gamma\mu$ along the balanced growth path.

Putting this all together in the arbitrage equation implies that

$$P_A = \frac{\pi}{r - n + \mu(1 - \gamma)} \tag{5.32}$$

is the price of a patent along the balanced growth path. One can see that this differs from the price of a patent in the Romer model in equation (5.18). Here, as the probability of a new innovation, μ, increases, the value of a patent declines. A higher probability of innovation means that the current capital good is more likely to be replaced quickly, making the value of the patent for the current capital good lower. Alternately, as the size of innovations, γ, increases the value of a patent increases.

5.4.4 SOLVING THE MODEL

We again have a model in which there are increasing returns in the aggregate production function, and the increasing returns require imperfect competition. Here, the imperfect competition shows up as the monopoly for the single intermediate-goods producer. These profits are extracted by the researchers who invent the new plans that allow a new intermediate-goods producer to replace the old intermediate-goods producer.

Note that it is always the case that a new innovation brings forth a new intermediate-goods firm. Why? This is due to the "Arrow replacement effect" of Kenneth Arrow (1962). The existing intermediate-goods firm will not bid as much for the patent of a new innovation, for while they will earn the profits from selling this new intermediate good, they will *lose* the existing profits they are earning. So to the existing intermediate-goods firm, new innovations are worth less than they are to a new firm. The new firm will always outbid the existing firm for the new patent, and it will replace them in supplying the intermediate good.

We already know the growth rate of the economy along the balanced growth path. What remains to solve for is the allocation of labor to research, s_R. As in the Romer model, we'll assume that individuals can work in the final-goods sector, earning.

$$w_Y = (1 - \alpha)\frac{Y}{L_Y}.$$

Alternatively, they could work as researchers, earning P_A if they innovate. They innovate with probability $\bar{\mu}$, so that their expected wage from research is

$$E[w_R] = \bar{\mu}P_A.$$

Unlike the Romer model, this is an expected wage, and the actual wage earned by a lone researcher is either zero (if he or she fails to innovate) or P_A (if he or she does innovate). By working in large groups, say at a research firm, researchers would be able to earn the expected wage rather than taking on the risk of innovation themselves. We'll assume that researchers are organized into large-scale research labs and earn precisely their expected wage.

With individuals free to move between research and working in the final-goods sector, it must be that $w_Y = E[w_R]$. As is shown in Appendix A, this can be used to solve for the share of the population that is engaged in research,

$$s_R = \frac{1}{1 + \frac{r - n + \mu(1 - \gamma)}{\alpha\mu}}. \tag{5.33}$$

This can be compared to the fraction of labor in research found in Section 5.2.4 in the Romer model, and they share a common component. The term $r - n$ appears in both, indicating again that the higher the discount rate that applies to profits, the lower the fraction working in research.

Looking further, we can see that there are two effects of μ. The first, in the term $r - n + \mu(1 - \gamma)$, represents the fact that as the chance of innovation increases, the value of a patent declines due to the higher probability of being replaced by the next innovator. In essence, this "business-stealing" effect of a higher μ causes innovators to discount the value of a patent more highly. This causes s_R to fall.[15]

The second effect, from the term $\alpha\mu$, represents the fact that if the probability of innovating goes up, then any individual researcher will be more likely to come up with an innovation and be able to sell the patent. Innovation becomes more lucrative, and so s_R rises. On net, what is the effect of an increase in μ? Innovations occur first, and only later are replaced. As individuals discount the future, the gains from innovation are large relative to the losses from replacement, and if μ increases then more people will work in the research sector. Mathematically, one can see this by taking the derivative of s_R with respect to μ.

COMPARING THE ROMER AND SCHUMPETERIAN MODELS

To a great extent, the two models of endogenous growth we've developed in this chapter provide identical results. For a realistic value of $\phi < 1$, the long-run growth rate is pinned down by the population growth rate n. So whether innovation takes the form of inventing entirely new

[15] This assumes that $\gamma < 1$. If γ were equal to one (or higher), then along the balanced growth path innovations would occur rarely, but when they did happen, they would double (or more) living standards. This seems unlikely to be a good description of the modern growth process.

intermediate goods or replacing existing intermediate goods is not essential to the long-run growth rate.

While the growth results are similar, a key contribution of the Schumpeterian approach is that it connects growth theory to the dynamics of firm behavior. For example, creative destruction means that new firms are entering and some existing firms are being destroyed. Recent research in growth, macro, trade, and industrial organization has used the Schumpeterian approach to explore a range of interesting issues, including the role of competition in promoting growth, firm dynamics, the direction of technical change, and the source of gains from exporting and international trade.[16]

The differences that arise between the models are in the *level* of income per capita, working through the share of labor engaged in research, s_R. Comparing equation (5.19) for the Romer model with equation (5.33) from the Schumpeterian model, you'll see that the exact solution for s_R differs slightly in the two models. Does one model of innovation imply a greater fraction of labor engaged in research? The answer is that it depends. The Schumpeterian model will have a higher s_R if $g_A < r - n$, or if the discount rate applied to profits is relatively large.[17] In this case, the future prospect of being replaced as the monopolist has little weight in an individual's evaluation of the gains from innovation, and so more people work at research. On the other hand, if the discount, rate $r - n$ is less than g_A, then individuals are particularly sensitive to the future "destruction" half of the creative destruction process and so do less research in the Schumpeterian world. In this case the Romer model will have a higher fraction of labor working in research, s_R.

Of course, in the real world the individuals engaged in research are a mix of those working on entirely new varieties and those attempting to creatively destroy an existing one and replace it. Regardless of whether we think the Romer model or Schumpeterian model is a better approximation of reality, the overall results that the long-run growth rate depends only on n, and that other policy changes have only level effects, hold with either style of innovation.

[16]For example, see Aghion, Bloom, Blundell, Griffith, and Howitt (2005), Acemoglu (2002), and Ramondo and Rodriguez-Clare (forthcoming).

[17]To see this, note that in the Schumpeterian model, $g_A = \gamma\mu$. We can write equation (5.33) as $[\gamma(r - n) + \gamma\mu(1 - \gamma)]/\alpha g_A$. This term will be smaller than $(r - n)/\alpha g_A$ from equation (5.19) if $r - n < \gamma\mu = g_A$. If this is true, then it must be that s_R is lower in the Schumpeterian model than in the Romer model.

5.5 OPTIMAL R&D

Is the share of the population that works in research optimal? In general, the answer to this question in both the Romer and Schumpeterian models is no. In each case, the markets do not induce the right amount of labor to work in research. Why not? Where does Adam Smith's invisible hand go wrong?

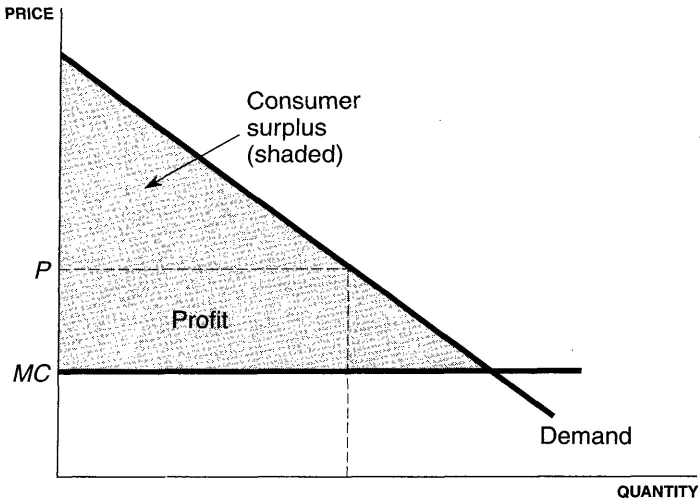
There are three distortions to research in the model that cause s_R to differ from its optimal level. Two of the distortions are easy to see from the production function for ideas. First, the market values research according to the stream of profits that are earned from the new design. What the market misses, though, is that the new invention may affect the productivity of future research. Recall that $\phi > 0$ implies that the productivity of research increases with the stock of ideas. The problem here is one of a missing market: researchers are not compensated for their contribution toward improving the productivity of future researchers. For example, subsequent generations did not reward Isaac Newton sufficiently for inventing calculus. Therefore, with $\phi > 0$, there is a tendency, other things being equal, for the market to provide too little research. This distortion is often called a "knowledge spillover" because some of the knowledge created "spills over" to other researchers. This is the "standing on shoulders" effect referred to earlier. In this sense, it is very much like a classic positive externality: if the bees that a farmer raises for honey provide an extra benefit to the community that the farmer doesn't capture (they pollinate the apple trees in the surrounding area), the market will underprovide honey bees.[18]

The second distortion, the "stepping on toes" effect, is also a classic externality. It occurs because researchers do not take into account the fact that they lower research productivity through duplication when λ is less than one. In this case, however, the externality is negative. Therefore, the market tends to provide too much research, other things being equal.

Finally, the third distortion can be called a "consumer-surplus effect." The intuition for this distortion is simple and can be seen by considering a standard monopoly problem, as in Figure 5.5. An inventor of a new design captures the monopoly profit shown in the figure. However, the potential gain to society from inventing the good is the

[18] On the other hand, if $\phi < 0$, then the reverse could be true.

FIGURE 5.5 THE "CONSUMER-SURPLUS EFFECT"



entire consumer-surplus triangle above the marginal cost (MC) of production. The incentive to innovate, the monopoly profit, is less than the gain to society, and this effect tends to generate too little innovation, other things being equal.

In practice, these distortions can be very large. Consider the consumer surplus associated with basic inventions such as the cure for malaria or cholera or the discovery of calculus. For these inventions, associated with "basic science," the knowledge spillovers and the consumer-surplus effects are generally felt to be so large that the government funds basic research in universities and research centers.

These distortions may also be important even for R&D undertaken by firms. Consider the consumer-surplus benefits from the invention of the telephone, electric power, the laser, and the transistor. A substantial literature in economics, led by Zvi Griliches, Edwin Mansfield, and others, seeks to estimate the "social" rate of return to research performed by firms. Griliches (1991) reviews this literature and finds social rates of return on the order of 40 to 60 percent, far exceeding private rates of return. As an empirical matter, this suggests that the positive externalities of research outweigh the negative externalities so that the market, even in the presence of the modern patent system, tends to provide too little research.

A final comment on imperfect competition and monopolies is in order. Classical economic theory argues that monopolies are bad for welfare and efficiency because they create "deadweight losses" in the economy. This reasoning underlies regulations designed to prevent firms from pricing above marginal cost. In contrast, the economics of ideas suggests that it is critical that firms be allowed to price above marginal cost. It is exactly this wedge that provides the profits that are the incentive for firms to innovate. In deciding antitrust issues, modern regulation of imperfect competition has to weigh the deadweight losses against the incentive to innovate.

SUMMARY

Technological progress is the engine of economic growth. In this chapter, we have endogenized the process by which technological change occurs. Instead of "manna from heaven," technological progress arises as individuals seek out new ideas in an effort to capture some of the social gain these new ideas generate in the form of profit. Better mousetraps get invented and marketed because people will pay a premium for a better way to catch mice.

In Chapter 4, we showed that the nonrivalrous nature of ideas implies that production is characterized by increasing returns to scale. In this chapter, this implication served to illustrate the general importance of scale in the economy. Specifically, the growth rate of world technology is tied to the growth rate of the population. A larger number of researchers can create a larger number of ideas, and it is this general principle that generates per capita growth.

As in the Solow model, comparative statics in this model (such as an increase in the investment rate or an increase in the share of the labor force engaged in R&D) generate *level effects* rather than long-run growth effects. For example, a government subsidy that increases the share of labor in research will typically increase the growth rate of the economy, but only temporarily, as the economy transits to a higher level of income.

The results of this chapter match up nicely with the historical evidence documented in Chapter 4. Consider broadly the history of economic growth in reverse chronological order. The Romer and Schumpeterian models are clearly meant to describe the evolution of technology since the establishment of intellectual property rights. It is the presence of patents and copyrights that enables inventors to earn profits to cover the initial costs of developing new ideas. At the same time, the world population was beginning to grow rapidly, providing both a larger market for ideas and a larger supply of potential innovators. In the last century (or two), the world economy has witnessed sustained, rapid growth in population, technology, and per capita income never before seen in history.

Consider how the model economy would behave in the absence of property rights. In this case, innovators would be unable to earn the profits that encourage them to undertake research in the first place, so that no research would take place. With no research, no new ideas would be created, technology would be constant, and there would be no per capita growth in the economy. Alternatively, consider world history with population fixed at the size found in 1 CE, roughly 230 million. Without growth in population, the economy is not taking full advantage of the increasing returns to scale that ideas provide. Even with property rights in place, the growth rate of technology would eventually fall to zero. Broadly speaking, a lack of property rights and a population growth rate close to zero prevailed prior to the Industrial Revolution.[19]

Finally, a large body of research suggests that social returns to innovation remain well above private returns. Although the "prizes" that the market offers to potential innovators are substantial, these prizes still fall far short of the total gain to society from innovations. This gap between social and private returns suggests that large gains are still available from the creation of new mechanisms designed to encourage research. Mechanisms like the patent system are themselves ideas, and there is no reason to think the best ideas have already been discovered.

[19]There were, of course, very notable scientific and technological advances before 1760, but these were intermittent and there was little sustained growth. What did occur might be attributed to individual curiosity, government rewards, or public funding (such as the prize for the chronometer and the support for astronomical observatories).

APPENDIX: SOLVING FOR THE R&D SHARE

ROMER MODEL

The share of the population that works in research, s_R, is obtained by setting the wage in the final-goods sector equal to the wage in research:

$$\bar{\theta} P_A = (1 - \alpha)\frac{Y}{L_Y}.$$

Substituting for P_A from equation (5.18),

$$\bar{\theta}\frac{\pi}{r - n} = (1 - \alpha)\frac{Y}{L_Y}.$$

Recall that π is proportional to Y/A in equation (5.14):

$$\frac{\bar{\theta}}{r - n}\alpha(1 - \alpha)\frac{Y}{A} = (1 - \alpha)\frac{Y}{L_Y}.$$

Several terms cancel, leaving

$$\frac{\alpha}{r - n}\frac{\bar{\theta}}{A} = \frac{1}{L_Y}.$$

Finally, notice that $\dot{A}/A = \bar{\theta} L_A/A$, so that $\bar{\theta}/A = g_A/L_A$ along a balanced growth path. With this substitution,

$$\frac{\alpha g_A}{r - n} = \frac{L_A}{L_Y}.$$

Notice that L_A/L_Y is just $s_R/(1 - s_R)$. Solving the equation for s_R then reveals

$$s_R = \frac{1}{1 + \frac{r - n}{\alpha g_A}},$$

as reported in equation (5.19).

SCHUMPETERIAN MODEL

The method to solve for s_R is similar to that used in the Romer model. First set the wage in the final-goods sector equal to the wage in the research sector:

$$\bar{\mu}P_A = (1 - \alpha)\frac{Y}{L_Y}.$$

Substitute in the value of patents from equation (5.32):

$$\bar{\mu}\frac{\pi}{r - n + \mu(1 - \gamma)} = (1 - \alpha)\frac{Y}{L_Y}.$$

From equation (5.30) we know that profits are proportional to Y, yielding

$$\frac{\bar{\mu}}{r - n + \mu(1 - \gamma)}\alpha(1 - \alpha)Y = (1 - \alpha)\frac{Y}{L_Y}.$$

Cancel common items and we have

$$\frac{\alpha}{r - n + \mu(1 - \gamma)}\bar{\mu} = \frac{1}{L_Y}$$

We defined the aggregate probability of innovation as $\mu = \bar{\mu}L_A$ in the text. Using this in the above equation gives us

$$\frac{\alpha\mu}{r - n + \mu(1 - \gamma)} = \frac{L_A}{L_Y}.$$

Again, $L_A/L_Y = s_R/(1 - s_R)$. Solving for s_R yields

$$s_R = \frac{1}{1 + \frac{r - n + \mu(1 - \gamma)}{\alpha\mu}},$$

which is what is shown in the text.

EXERCISES

1. *An increase in the productivity of research.* Suppose there is a one-time increase in the productivity of research, represented by an increase in θ in Figure 5.1. What happens to the growth rate and the level of technology over time?

2. *Too much of a good thing?* Consider the level of per capita income along a balanced growth path given by equation (5.11). Find the value for s_R that maximizes output per worker along a balanced growth path for this example. Why is it possible to do too much R&D according to this criterion?

3. *The future of economic growth* (from Jones 2002). Recall from Figure 4.6 and the discussion surrounding this figure in Chapter 4 that the number of scientists and engineers engaged in R&D has been growing faster than the rate of population growth in the advanced economies of the world. To take some plausible numbers, assume population growth is 1 percent and the growth rate of researchers is 3 percent per year. Assume that $\dot{A}/A$ has been constant at about 2 percent per year.

   (a) Using equation (5.6), calculate an estimate of $\lambda/(1 - \phi)$.

   (b) Using this estimate and equation (5.7), calculate an estimate of the long-run steady-state growth rate of the world economy.

   (c) Why does your estimate of long-run steady-state growth differ from the 2 percent rate of growth of A observed historically?

   (d) Does the fact that many developing countries are starting to engage in R&D change this calculation?

4. *The share of the surplus appropriated by inventors* (from Kremer 1998). In Figure 5.5, find the ratio of the profit captured by the monopolist to the total potential consumer surplus available if the good were priced at marginal cost. Assume that marginal cost is constant at c and the demand curve is linear: $Q = a - bP$, where a, b, and c are positive constants with $a - bc > 0$.