5 THEORIES OF ENDOGENOUS GROWTH

Chapter 5 Contents

A. Topics and Tools

The neoclassical growth theory that we studied in Chapters 3 and 4 largely evolved in the 1950s. There was considerable filling-in of details in the 1960s, but by the 1970s growth theory had largely become moribund. A tremendous revitalization has occurred since the 1980s, spurred by several shortcomings of the previous theories.

First, because growth rates are exogenous in the Solow and Ramsey models, these theories are unable to explain why growth rates (and, in particular, the rate of technological progress) might change from one time period to another. This became an important research topic in the 1980s when emerging data began to convince macroeconomists that productivity growth in the United States and other advanced countries had declined significantly beginning about 1974.

A second failing of neoclassical growth theory is that it cannot explain the large and lasting differentials in per-capita income that we observe across countries and regions. Solow's growth model implies more rapid convergence of incomes than seems actually to have occurred, particularly between developed and developing countries. International differences in technological capability can help explain this gap, but beg for an economic explanation that cannot be provided by models in which technology is exogenous.

Another feature of neoclassical growth models that some economists and policymakers find troublesome is that they provide no mechanism by which the saving and investment rate (or government policies directed at influencing it) can affect the steady-state growth rate. While this conclusion of neoclassical models is not *obviously* counterfactual, many find it counterintuitive and have explored models in which saving plays a more central role.

The pioneer of "endogenous growth theory" is Paul Romer, a former colleague but not a relative of our textbook author.[1] His 1986 paper in the *Journal of Political Economy* is a seminal work in the modern revitalization of growth theory. The principal engine behind endogenous growth is the elimination of the assumption of decreasing returns to "capital."[2] In order to justify this radical departure from a long-

---

[1] In the early 1990s, there were *three* famous young Romers teaching macroeconomics at the University of California at Berkeley. Paul, who focuses on growth theory and is now at the Stanford Business School, David (our author), who is a prominent neo-Keynesian, and David's wife Christina, who is a macroeconomic historian and now chair of the Council of Economic Advisors in the Obama Administration.

[2] This is a good time to clarify two closely related concepts: "diminishing marginal returns" and "decreasing returns to scale." The former is usually applied to changes in only a single factor of production holding all other factors constant. Thus, diminishing returns to capital means that when more capital is added to production with all other factors held constant, the ensuing increase in output becomes smaller as more and more capital is added. Returns to scale usually apply to the effect on output of simultaneous changes in many or all factors of production. "Constant returns to scale" by itself means that increases of an equal percentage in all factors leads to an increase of the same percentage in output. In this chapter, we will extend the idea of returns to scale to situations where a subset of factors changes.

established assumption of microeconomic theory, Romer and his followers have broadened the definition of capital to include human capital and/or knowledge capital. As we shall see, once this broader view of capital is adopted it is no longer obvious that there are decreasing returns. This leads to radical changes in the conclusions that we derive from models that are otherwise similar to those of Solow and Ramsey.

There are three basic models developed in the chapter: the R&D model of 3.2 and 3.3, the human-capital model of 3.8, and the production-protection-predation model of section 3.11. We will give attention to all three.

The mathematical tools used here are largely familiar ones. To keep the analysis simple, Romer reverts to the simple Solow assumptions about saving (and other static resource allocation decisions). The original literature on these models bases decisions on utility and profit maximization, which is more satisfactory, but the dynamic properties of the model are similar with constant growth rates, so Chapter 3 will teach you the essential features of the model without all the complicated mathematics that we saw in Romer's Chapter 2.[3]

As in previous chapters, we will be searching for steady-state balanced growth paths. To find these, we will usually look for situations in which the growth rates of the key state variables are constant. In most of the models of this chapter, there will be two state variables, either physical capital and knowledge capital or physical capital and human capital. We will use a two-dimensional phase plane that looks on the surface like the one in the Ramsey model, but is fundamentally different because in these models both variables are state variables that cannot jump, whereas in the Ramsey model c was a control variable that could jump vertically to adjust to changes in economic conditions.

B. The Microeconomics of Innovation and Human Capital Investment

Romer's Chapter 3 examines the *macroeconomic* implications of investment in research and development (innovation) and human capital. However, some of the most important theoretical issues in modeling these concepts are *microeconomic* in nature. The seminal papers in the modern growth literature vary a lot in how careful-

[3] Barro and Sala-i-Martin (2004) is a more advanced textbook that looks at more sophisticated versions of these models. Acemoglu (2009) is a more recent, and more mathematical, treatment. For those interested in learning about them, Econ 454 develops the more complete models.

ly they model these microeconomic issues, but Romer's simplified presentation of the models largely ignores the microeconomics. In this section, we briefly consider some of the basic microeconomic issues involved. (Romer discusses some of these topics in Section 3.4.)

The models of Chapter 3 attempt to make endogenous the "production" of technology. In the R&D model, an R&D sector produces additions to society's stock of technical knowledge. In the second, individuals add to their human capital by spending time in education rather than producing output.

A key microeconomic issue that underlies this analysis is the question of what incentive people have to make investments in knowledge or in human capital. Unless people get utility directly from the process of research or education (which cannot be ruled out—consider the case of the "professional student"), they will only undertake these investments if they are able to profit from them sufficiently to justify the opportunity cost. The opportunity cost of investing in research or education may include both forgone consumption and the lost alternative opportunity of investing in physical capital. Thus, if rational agents invest in research or education, then the earnings from these activities must have an expected *present value* at least as high as the current consumption that must be forgone in order to undertake them *and* as high as the expected present value of the returns to physical capital investment, since that is also an alternative use of funds.

Returns to research and development

As Romer discusses on page 116, pure knowledge is ***nonrival***, meaning that the use of knowledge by one person does not reduce the ability of others to use it. Most "private" goods in the economy are, by contrast, rival. To clarify the distinction, think about chocolate-chip cookies.[4] Everyone can use the same (non-rival) recipe for chocolate-chip cookies but everyone cannot use the same (rival) chocolate chips.

As you learned in Econ 201, a competitive market economy (in the absence of externalities) can lead to the production of the efficient amount of traditional, rival goods. The market price provides producers and consumers with a scarcity signal that can lead to efficient resource allocation by equating the marginal social cost of the good with its marginal social benefit. On the production side, producers equate price to the marginal production cost. Consumers consume at the level where the marginal benefit of an additional unit of the good equals the market price.

Nonrival goods such as knowledge can be reused by the same person or shared with additional people at *zero* marginal social cost. With marginal cost equal to zero, efficiency requires that people should consume knowledge at the level where its marginal social benefit is also zero. But, as with rival goods, utility-maximizing or profit-

---

[4] Some of us will use any excuse to justify thinking about chocolate-chip cookies.

maximizing users of knowledge will "purchase" it up to the point its marginal benefit equals the price that is charged. They will voluntarily choose the *optimal* level of use (where marginal benefit is zero) only if the price of knowledge is zero. Thus efficiency requires that knowledge, once it has been created, must be distributed freely at a zero price.

However, if the market price of knowledge is zero, then the market provides no financial reward for anyone who incurs the research-and-development costs that are necessary to create it. To provide such incentives, most countries have patent and copyright laws that grant exclusive (monopoly) intellectual property rights to individuals who create knowledge. With a patent or copyright, the creator is able to charge a positive royalty for the license to use knowledge or to earn monopoly profits on the newly discovered product or process by using it exclusively and prohibiting its use by others. However, the argument of the preceding paragraph shows that charging a positive price for the use of a nonrival good such as knowledge leads to inefficiency, as does the existence of patent-protected monopolies. If individuals must pay to use knowledge, but the social cost of it using it is zero, then they will choose to use knowledge at a lower-than-optimal level. Thus, when intellectual property laws work as intended, they help resolve one problem by encouraging investment in knowledge, but at the same time they create another by discouraging its use. This argument is familiar to anyone who has followed the sometimes intense debates about illegal copying of software, works of entertainment such as CDs and DVDs, or about the pricing of pharmaceuticals.

While patents provide substantial protection in some industries, such as chemicals, intellectual-property laws often do not give innovators much protection. For example, suppose that a company discovers that a particular tool works better if it is curved than if it is straight. It can attempt to profit from its discovery by patenting the curved tool. However, there are many ways to curve a tool and it is probably impossible to gain patent rights on *all* possible curves that might be beneficially used. Once the knowledge that curved tools are better becomes public (as it does when a patent issues), everyone may be able to adopt some variant of the improved technology without paying a royalty to the inventor. For such ***nonexcludable*** kinds of knowledge, inventors often resort to secrecy in hopes that it will be costly and time-consuming for competitors to discover or "reverse engineer" the knowledge. When knowledge is both nonrival and nonexcludable, it qualifies as a pure public good, with all the familiar resource-allocation problems that public goods entail. Governments often subsidize research and development for branches of knowledge where nonexcludability makes patent protection ineffective or where wide diffusion of the resulting knowledge seems especially important.

The issue of the efficient allocation of resources to research and development is a central focus of Reed's Economics 354: The Economics of Science and Technology.

If you are interested in pursuing additional readings in this area, visit the instructor's Web page for a link to a recent reading list.

Human capital vs. knowledge capital

By human capital we mean acquired characteristics that make workers more productive. Although it encompasses such characteristics as health, strength, and stamina, the most commonly analyzed sources of human capital are the education, training, and experience that a worker embodies. Since education and training involve the transmission of knowledge, it might seem like human capital is the same as the knowledge capital we study in the R&D model.

However, there is a crucial difference. Knowledge capital is potentially a public good whereas human capital is not. Perhaps the easiest way of distinguishing between them is to think about the two major roles that most professors play. You see professors most often in the classroom, where they are imparting existing knowledge to students. This increases the students' human capital, but does not create new knowledge for society. When they are not in the classroom, your professors are likely to be engaged in research. If successful, this research leads to new knowledge capital that everyone can potentially share on a nonrival basis. Thus, simply put, society's knowledge capital is everything that is known by someone in the society; your human capital includes your personal familiarity with and ability to use part of that knowledge. Your human capital is personal to you—the fact that you have obtained knowledge may make you more productive but it does not usually raise anyone else's productivity. Thus human capital does not have the public-good characteristics of knowledge capital.

Returns to education

Although human capital is not a public good in the same way as knowledge capital, education raises interesting economic issues of its own. Some aspects of education have elements of nonrivalry. The syllabus for a course or a recorded lecture can be shared widely at minimal cost. However, most other aspects of education are rival. Classroom seats and instructor time are limited and putting one student into a seat denies that seat to someone else. Moreover, most kinds of education are easily excludable. Those who do not pay for a seat in the class can be denied access. Thus, it does not appear that education is a public good.[5]

One can imagine an uncomplicated world in which markets could allocate education efficiently. If the benefits of a person's education and training (including on-

---

[5]Some economists argue that everyone gains from having a more educated society, so that additional education for one individual benefits others as well as himself. In this case, there is a positive externality and the market system of incentives will lead to underconsumption of education.

the-job training) are perfectly reflected in his or her enhanced productivity, then someone who has acquired more human capital should receive commensurately higher wages. In this case, the individual can make an optimal personal decision about whether the returns to further human-capital acquisition (higher productivity and wages) justify the cost. There is no market failure here and education/human capital is similar to other kinds of investment/capital.

However, there are several problems that may upset efficient resource allocation in education markets. One is the problem of borrowing to finance education investment. Investments in capital—whether physical, human, or "knowledge capital" acquired through research and development—require a substantial initial expenditure, followed by a lengthy period over which the investment earns a return. The person desiring to make the investment often does not have sufficient liquid funds at the time the investment is to be made, so a "capital market" in which one can borrow for such expenditures is a useful social institution. However, capital markets can only function if lenders can be reasonable sure that they will be repaid.

A borrower who purchases a physical capital good such as a building or a machine must normally pledge the capital good as collateral on the loan. If the borrower fails to repay the loan as required, the lender can seize the capital good and resell it to recover at least part of his or her money. However, a borrower's education cannot be seized and, in societies that outlaw slavery, borrowers themselves cannot be seized by the lender either.[6] This limits the recourse of lenders in cases of default, which makes it hard for the private market to provide access to loans for human-capital investment.

Government-subsidized student loans attempt to remedy this market failure by providing government guarantees in place of collateral. Many Reed students can confirm that this allows a thriving market in student loans, but it does not assure allocative efficiency. Government guarantees generally make student loans less risky to lenders than the intrinsic economic riskiness of the underlying returns to education. Thus, interest rates will usually be subsidized below the level that would be appropriate to the investment's risk and this will encourage overuse of student loans. Furthermore, while the government guarantees allow the market to function, the government is often no more effective at collecting money from defaulters than a private lender. This diverts the cost of human capital from the investor/student onto the general taxpayer.

A second difficulty in human-capital investment arises when human capital is acquired through on-the-job learning. In most jobs, the worker learns a great deal about how the job is done during the initial months of employment. During that period, productivity increases rapidly as the worker gets better at what he or she does.

[6] Notwithstanding the fellow who was "repossessed" after failing to pay his exorcist's bill.

A "perfect" market might capture this learning by starting the individual at an extremely low wage (or the new worker might even pay the firm for the privilege of learning the job), then increasing the wage as productivity rises. This scheme implicitly or explicitly makes the worker pay for the investment in human capital. To the extent that workers value the human capital they acquire, they may be willing to incur this cost, although if the initial wage is low enough it might force them into a borrowing situation that raises the same problem of collateral described above.

However, much of the knowledge acquired on the job may be "firm-specific" human capital, such as knowledge of the internal rules and operations of a particular organization, and be largely useless if a worker moves to another firm. In a world in which layoffs and job changes are common, workers will be reluctant to bear the cost of training that is useful *only* when he or she works for one particular employer. The firm will also hesitate to invest in a particular worker when the worker might depart, though there is less risk of the worker quitting at the end of the training period if most of the training is firm-specific. These difficulties in appropriating the returns to human-capital investment can lead to underinvestment in training.

A third problem that complicates the efficient allocation of resources to human capital is that the link between education and productivity is not well understood. It is uncontroversial that more highly educated workers are more productive; what is at issue is whether people who are innately more productive tend to invest in more education or whether it is the education itself that makes them productive. In our growth models, we assume that education makes individuals more productive. However, some economists argue that education acts mostly as a screening or ranking device. According to this "signaling" theory, firms hire college graduates at high wages not because they have learned anything that makes them more productive, but because the fact that they finished college signals that they are individuals of high ability and potentially high productivity.

If one takes this signaling argument to its extreme, then one may claim that education has little effect on productivity; it just acts like an elaborate placement test for employers. For example, a century ago a relatively small share of people finished high school and very few finished four years of college. According to the signaling model, being a high-school graduate at that time signaled that you were a high-quality worker and being a college graduate signaled that you were in an elite category of high achievers. Today the majority of people finish high school, so the signaling value of a high-school diploma is small. Many individuals finish college, so even a college degree is no indicator of exceptional ability. To demonstrate a truly elite status one must now attend graduate school and get an advanced degree. According to the extreme version of the signaling theory, the people who now get good jobs with a graduate degree used to be able to get the same jobs (and do them just as well) with a bachelor's degree. If the additional years of study do not raise productivity, then they are a costly waste of resources.

Although most economists believe that education makes individuals more productive, it is difficult to disprove the signaling model because in many cases the two models lead to similar predicted outcomes. The human-capital model in Romer's Chapter 3 assumes that education is an investment in human capital that enhances workers' productivity. However, to the extent that education is mainly a signaling tool, this model may overstate the benefits of education.

C. Understanding Romer's Chapter 3, Part A

Introduction

Chapter 3 examines several strains of modern research literature on economic growth. The first approach models the production of improvements in technology by including "knowledge capital" along with physical capital. A two-sector model is required because knowledge production does not follow the same production function as goods production; there is an R&D (or knowledge production) sector alongside the usual sector producing physical goods.

The introduction of a second sector requires the use of some new modeling techniques. For example, aggregate resources must now be divided between the production of "goods"—either physical capital or consumption goods—in one sector and the production of knowledge or human capital in the other. This is the role of the a coefficients in Romer's R&D model.

The crucial novelty of these models that makes their conclusions strongly different than the ones of the earlier chapters is that the introduction of human or knowledge capital may allow us to sidestep the usual assumptions of diminishing returns to capital and constant overall returns to scale. While it is intuitively clear that adding more physical capital to a given amount of labor must eventually lead to a diminishing marginal product of capital, there is no obvious reason why increases in knowledge would be subject to such diminishing returns. Moreover, knowledge spillovers from one producer to another may allow increasing returns to scale for the economy as a whole, even if traditional factors (labor and physical capital) produce with constant returns to scale for any given state of technology.

Romer's modeling strategy and the research literature

As Romer notes early in Part A, the model he presents is a simplified version of a family of models that evolved in the growth literature in the early 1990s. A look at the research papers he cites on pages 101 and 102 will verify for you that he has made several simplifying assumptions. What he has done is to describe in detail a simple version of this class of models that preserves their essential features.

For example, Romer's model uses a Cobb-Douglas production function. This allows us to evaluate marginal products explicitly and solve growth-rate equations that would otherwise have only implicit solutions. Similarly, he relies on the assumption of a constant saving rate in most of Chapter 3 rather than building utility maximization into the model, though this is relaxed in section 3.5. The more general models lead to the same qualitative conclusions, so we have gained expositional simplicity without losing the basic logic of the model. Students who are interested in the more general approach are strongly encouraged to explore the papers cited in Chapter 3 or to take Economics 454: Economic Growth, in which these models are examined in more detail.

The basic setup of the R&D model

A key difference from the previous growth models that you have encountered is that this model has two sectors. There are two kinds of capital in the models of Chapter 3. In the R&D model, there is physical capital, which is familiar from earlier models, and "knowledge capital." Since there are two stocks, or ***state variables***, we have two equations of motion and must analyze convergence jointly.[7] Romer builds up to this gradually by first ignoring physical capital and looking at the implications of knowledge investment as a single state variable (in Section 3.2). He then brings physical capital back in to create the formal two-state-variable model in Section 3.3.

With two "produced" or "reproducible" factors, an additional dimension of choice is available to the economy. Individuals can use their labor and capital resources either in the sector producing physical goods or in the sector producing knowledge. How much of the economy's resources will be dedicated to producing knowledge rather than goods? This is a complicated question for several reasons. As discussed in the previous section of this chapter, there are significant conceptual differences between these kinds of capital that may have important implications for the

[7]This is a lot like what we did with c and k in the Ramsey model, but it is a little different. In the Ramsey model, c is a "control" variable rather than a state variable because it can jump discretely at an instant of time. (If something happens to change their situation, consumers can raise or lower the flow of consumption spending at time t in response.) This was crucially important in allowing the model to converge along the saddle path. If c did not jump exactly to the value required by the saddle path, the model would have been unstable. Stock or "state" variables such as capital and knowledge cannot jump in the same way. Their value at any instant depends only on past investment; they change smoothly through the equations of motion of the model. Although we can imagine discrete jumps in these variables—perhaps a disaster that destroys capital instantly—this would imply a momentary suspension of the equation of motion that says that depreciation is proportional to the stock.

incentives of the private sector to invest in them. Romer avoids this issue in the construction of the model by assuming that a_K and a_L, the shares of capital and labor devoted to knowledge production, are exogenous. A more satisfactory approach (that is taken in most of the research literature) would be to endogenize these values by examining the markets for factors of production in detail and modeling the choice of owners of factors about the industry to which they sell their resources.

The knowledge production function

There are several features of knowledge production that are worth stressing in this model. First, knowledge does not depreciate. From Romer's equation (3.2), it is clear that if a_K and a_L are zero (so that no resources are devoted to knowledge accumulation), the stock stays constant, $\dot{A}(t) = 0$.

The absence of depreciation of knowledge may seem counterintuitive, since old knowledge does not seem to be worth very much in today's world. However, we must distinguish between the *usefulness* of a specific nugget of knowledge and the *existence* of the nugget of knowledge itself. The usefulness may decline even if we do not have depreciation of the aggregate stock itself. The knowledge of how to produce 1980-vintage computers is only useless today because it has been superseded by even more modern knowledge (most of which builds on the original knowledge). A reasonable way to think about the absence of aggregate depreciation of knowledge in equation (3.2) is that technical knowledge does not disappear or wear out with use (like physical machines do). An economy that devotes no resources to the production of knowledge does not slide backwards; it merely fails to progress.[8]

A second feature of the knowledge production function is the possibility of increasing or decreasing returns to scale in its production. As Romer notes on page 102, the usual "replication" argument for aggregate constant returns to scale does not apply to the production of knowledge. He presents reasons why returns to scale might be either decreasing or increasing.

Finally, the role of the θ parameter in equation (3.2) is very important. (This θ is totally unrelated to the θ in the CRRA utility function that we used in the previous chapter.) To see the intuition of the role that θ plays in the analysis, divide both sides of (3.2) by $A(t)$ to get

$$\frac{\dot{A}(t)}{A(t)} = B[a_K K(t)]^{\beta}[a_L L(t)]^{\gamma} A(t)^{\theta-1}. \tag{1}$$

[8] There have been historical instances of knowledge being lost. Before written archives of technological literature, knowledge of techniques could die with the individuals who knew them. Landes (1983) describes the example of the Chinese water clock constructed by Su Sung in 1094.

Equation (1) expresses the knowledge production function in terms of the growth rate of technology (in percentage terms). Suppose that the amounts of capital and labor allocated to knowledge production are fixed, *i.e.*, $a_K K(t)$ and $a_L L(t)$ are constant. Equation (1) shows that this will lead to a constant rate of technical progress if $\theta = 1$, since $A(t)^0 = 1$ and thus $A(t)$ vanishes from the right-hand side. This corresponds to the kind of progress assumed in the Solow and Ramsey models: growth in $A(t)$ at a constant rate g.[9]

If $\theta > 1$, then $\theta - 1 > 0$ and an increase in $A(t)$ will cause the growth rate of $A(t)$ to increase, given fixed amounts of capital and labor devoted to knowledge production. In other words, when $\theta > 1$, the more knowledge we have, the faster the stock of knowledge grows for a given amount of resources devoted to knowledge production. More knowledge accelerates the growth rate, which of course raises the level of knowledge even more rapidly, causing a further increase in the growth rate, and so on. Not surprisingly, this condition turns out to be associated with explosively accelerating growth in technology, productivity, and output.

If $\theta < 1$, then $\theta - 1 < 0$ and each increase in $A(t)$ lowers the growth rate of $A(t)$, other factor inputs held constant. In this case, technology exhibits a kind of diminishing returns with respect to its own production that is similar to that of capital in the Solow model.

Capital-generated growth in the Solow model was limited by the fact that capital faced diminishing returns in reproducing itself. Eventually, the economy settled into a steady state in which capital could no longer grow relative to other factors. The same thing happens to technology in the R&D model when $\theta < 1$: Eventually, technology-induced growth is limited and, without growth in the non-produced factor (labor), the economy becomes stationary (zero growth).

Analysis of the model without capital

Romer uses phase diagrams to search for a steady state in this model, just as we did in the Solow model when we plotted $\dot{k}$ as a function of k and looked for the point at which the curve intersected the horizontal axis. However, in the Solow analysis, we were looking for a "stationary value" of k, which was a ratio among variables in the model (K/AL). The analysis represented in Romer's Figure 3.1 is similar, but here we seek a stationary value for g_A, the growth rate of technology, instead of k.

---

[9] However, note that in a Solow/Ramsey steady state the total quantities of labor and capital are increasing, so if the a values are constant $a_K K(t)$ and $a_L L(t)$ would be increasing over time and $\theta = 1$ does not automatically lead to an equilibrium growth path similar to those models.

To find a stationary value of g_A in the R&D model, we need to examine $\dot{g}_A$, the change over time in the growth rate of $A(t)$. That means that we are looking at the change in a growth rate, which might be a little confusing at first.

Note that the sign of g_A (and *not* the sign of $\dot{g}_A$) tells us whether $A(t)$ is growing or shrinking. If g_A is positive, then A is growing; if $g_A < 0$ then it is shrinking. The condition that $\dot{g}_A > 0$ means that the *growth rate* g_A (whether positive or negative) is getting larger as time passes. Similarly, the statement that $\dot{g}_A < 0$ means that the *growth rate* of $A(t)$ is getting smaller as time passes.

The intermediate case of $\dot{g}_A = 0$ is the case where the growth rate of $A(t)$ is constant. This situation could be a steady-state, constant-growth equilibrium. Thus, our search for a steady state involves finding conditions under which $\dot{g}_A = 0$, then assessing whether the economy would converge to such a state.

In order to find the steady state in which $\dot{g}_A = 0$, Romer first derives an expression for $g_{A,}$ which is just our equation (1) with β set to 0 to reflect the no-capital assumption (Romer's equation (3.7)). To get an expression for $\dot{g}_A$, he takes the derivative of (3.7) with respect to time to get (3.8). Since getting (3.8) from (3.7) is not obvious upon inspection, let's examine the intervening steps.

Looking closely at (3.7), Ba_L^{γ} is constant over time, so it will not play an important role in the time derivative. $L(t)$ and $A(t)$ both vary with respect to time, and their powers are multiplied by each other in (3.7), so we will need to use the product rule and the chain rule to differentiate with respect to time. Applying these rules directly to (3.7) yields

$$
\begin{aligned}
\dot{g}_A(t) &= Ba_L^{\gamma}\left[\gamma L(t)^{\gamma-1}A(t)^{\theta-1}\dot{L}(t) + (\theta-1)L(t)^{\gamma}A(t)^{\theta-2}\dot{A}(t)\right] \\
&= \gamma Ba_L^{\gamma}L(t)^{\gamma}A(t)^{\theta-1}\frac{\dot{L}(t)}{L(t)} + (\theta-1)Ba_L^{\gamma}L(t)^{\gamma}A(t)^{\theta-1}\frac{\dot{A}(t)}{A(t)} \\
&= \gamma g_A(t)\frac{\dot{L}(t)}{L(t)} + (\theta-1)g_A(t)\frac{\dot{A}(t)}{A(t)},
\end{aligned}
$$

which simplifies to

$$
\dot{g}_A(t) = \left[\gamma n + (\theta-1)g_A(t)\right]g_A(t). \tag{2}
$$

Equation (2) can be rewritten as Romer's equation (3.9):

$$
\dot{g}_A(t) = \gamma n g_A(t) + (\theta-1)[g_A(t)]^2,
$$

which shows that $\dot{g}_A(t)$ is a quadratic function of $g_A(t)$. That means that we get a parabola if we graph $\dot{g}_A(t)$ as a function of $g_A(t)$. We can use basic algebra to examine the characteristics of this parabola. Because there is no constant term in this quadratic function, $\dot{g}_A(t) = 0$ when $g_A(t) = 0$ and it the parabola must pass through the origin. The slope of the function is its derivative with respect to $g_A(t)$,

$$\frac{d\dot{g}_A(t)}{dg_A(t)} = \gamma n + 2(\theta - 1)g_A(t). \tag{3}$$

At the origin, $g_A(t) = 0$, so the slope expression of equation (3) is $\gamma n > 0$ and the function is sloping upward at the origin.

The sign of the coefficient on the squared term, which is $\theta - 1$ in (2), determines the convexity or concavity of the parabola. If $\theta - 1 > 0$ then the slope of the function is increasing from left to right and the parabola opens upward. Since it starts at the origin with a positive slope, this means that it heads upward at an increasing rate as shown in Romer's Figure 3.2. If $\theta - 1 < 0$, then the slope is decreasing and the parabola opens downward, reaching a maximum in the positive quadrant and intersecting the horizontal axis as shown in Romer's Figure 3.1. In the borderline case where $\theta - 1 = 0$, the function is a positively sloped straight line coming out of the origin, as in Romer's Figure 3.3. (The straight line is a special case of the parabola in which the coefficient on the squared term is zero.)

Clearly, the decisive condition determining the shape of the parabola (and therefore the dynamic behavior of $g_A(t)$) is whether θ is greater than, less than, or equal to one. This provides mathematical support for our discussion in the previous section, where the magnitude of θ was asserted to be very important.

When $\theta < 1$, growth in technology is not "self-sustaining" due to diminishing returns to knowledge. Past discoveries make future discoveries more costly in terms of resources. Positive technological progress can only be sustained in this case if growth in the labor force allows more and more labor resources to be devoted to research as time passes. Note that if $n \leq 0$, then the slope of the $\dot{g}_A(t)$ function is zero or negative at the origin. With $n \leq 0$ and $\theta < 1$, the $\dot{g}_A(t)$ function immediately turns downward into the negative quadrant. In this case, the economy has steadily decelerating technical progress approaching a steady state in which $g_A(t) = 0$. Thus, we conclude that in the case where $\theta < 1$, only steady growth in the labor force will allow positive technological progress in a steady state.

If $\theta > 1$, the rate of technological progress may grow explosively. Each discovery opens up multiplying new opportunities so that future discoveries become less costly to find. Progress feeds on itself so strongly that growth in technology can accelerate endlessly even with constant resources devoted to R&D. If $n \geq 0$, there is no point to the right of the origin at which the curve intersects the horizontal axis, so there is no nonzero steady-state rate of technical progress.

It was noted in the previous section that if $\theta = 1$, then the growth rate of technology is neither enhanced nor retarded by the pre-existing level of technology. If the labor force does not grow ($n = 0$) and $\theta = 1$, then technological progress will occur at a constant rate. Both terms on the right of equation (3) are zero, so $\dot{g}_A(t) = 0$ and $g_A(t)$ remains at $Ba_L^{\gamma}L^{\gamma}$, the level dictated by Romer's equation (3.7). (Note that L does not require a time index since it is constant when $n = 0$.) In this case, the line in Romer's Figure 3.3 coincides with the horizontal axis, meaning that *any* level of technology growth seems to be a potential steady state—whatever the growth rate of A, it will remain constant. The technology production function tells us that the growth rate that the economy starts and remains at is $Ba_L^{\gamma}L^{\gamma}$.

There are several key characteristics of the model with $\theta = 1$ that make it interesting to growth economists. First, this is a case where increasing the allocation of resources to research (a_L) leads to a higher steady-state growth rate. That is the sense in which models of this kind are called "endogenous" growth models. An economy that makes an economic choice to devote more of its resources to accumulating knowledge capital (perhaps through a policy of subsidizing R&D) will have a permanently higher growth rate. By contrast, the Solow model predicts that economies that devote more resources to capital accumulation (saving) will have a higher level of income, but not a permanently higher growth rate. Thus, changes in the rate of capital investment have "growth effects" in endogenous-growth models but just "level effects" in convergent models such as Solow's.

Second, if we change our assumption about growth in the labor force to allow $n > 0$, then increased labor input over time will result in everlasting acceleration of technological progress in the $\theta = 1$ case. A growing population means (for given a_L) more scientists, which means more discoveries and faster technological advance. Since knowledge is assumed to be nonrival, each discovery is costlessly shared by all, so it is the total amount of knowledge created that drives growth, not knowledge-creation per capita.[10]

Finally, as Romer notes on page 108, the crucial parameter in determining the dynamics of the system is the magnitude of returns to scale *to produced factors*. By this we mean "Does a doubling of only the produced factors leads to a doubling or more or less than a doubling of production?" The Solow model had constant returns to scale to *all* factors (labor and capital), but diminishing returns to the single *produced*

[10] Some growth models have made the alternative assumption: that knowledge is strictly private. In these cases, it is knowledge production per capita that matter for growth. In Chapter 6 we consider a paper by Peter Klenow (1998) that tests this assumption empirically.

factor (capital). Diminishing returns to produced factors assure that the $sf(k)$ curve in the Solow model is convex, making convergence to a steady state inevitable and ruling out self-sustaining "endogenous" growth in the capital stock.

Since knowledge is the only produced factor in the R&D model of section 3.2, the relevant condition for returns to scale in produced factors is whether θ is greater than, less than, or equal to one. This is exactly the condition that we showed above to have a decisive effect on the dynamic properties of the model. We shall find that this is a quite general proposition in this class of models: growth is self-limiting, self-sustaining, or explosive depending on whether returns to scale to produced factors are decreasing, constant, or increasing.

The R&D model with capital

As noted above, the biggest methodological difference between the full R&D model and previous models is the presence a second state variable K along with A. We now explore the full, two-state-variable version of the R&D model that Romer presents in section 3.3.

To search for a steady state, we now seek a point at which the growth rates of *both* state variables are constant over time. In other words, in addition to seeking conditions under which $\dot{g}_A(t) = 0$, we must also find conditions that lead to $\dot{g}_K(t) = 0$.

Since $\dot{g}_A(t)$ and $\dot{g}_K(t)$ will, in general, both depend on the current values of both $g_A(t)$ and $g_K(t)$, we will have to use a two-dimensional phase diagram. Romer's Figures 3.4 through 3.7 build such diagrams for two cases of the model. As we did for the Ramsey model, we divide the space of possible values for g_A and g_K into regions according to whether $\dot{g}_A$ and $\dot{g}_K$ are respectively positive or negative. To do this, we plot the curves corresponding to the conditions $\dot{g}_A = 0$ and $\dot{g}_K = 0$. We then use arrows to indicate the directions of horizontal and vertical motion from any point.

For the general R&D model, it turns out that both of the relevant curves are upward-sloping lines. The line corresponding to $\dot{g}_K = 0$ has a positive vertical intercept and a slope of one; the line for $\dot{g}_A = 0$ has a negative vertical intercept and a slope of $(1 - \theta)/\beta$. The behavior of the system depends on the relative slopes of the two lines: whether $(1 - \theta)/\beta$ is greater than, less than, or equal to one.

The $\dot{g}_A = 0$ line starts below the $\dot{g}_K = 0$ line, since the former has a negative intercept and the latter a positive one. If $(1 - \theta)/\beta > 1$, then the $\dot{g}_A = 0$ line has a steeper slope and will eventually intersect the $\dot{g}_K = 0$ line. Thus, if $(1 - \theta)/\beta > 1$, the model has a unique steady state with the growth rates of capital and technology settling down to constant values. Alternatively, if $(1 - \theta)/\beta = 1$, the lines are parallel and if $(1 - \theta)/\beta > 1$, the lines not only never intersect in the positive quadrant but are getting farther apart as the economy moves away from the origin. In these cases, there is no unique steady state.

The dynamic character of the model thus depends on the magnitude of $(1 - \theta)/\beta$ relative to one. Note that $(1 - \theta)/\beta = 1$ if and only if $1 - \theta = \beta$, or $\beta + \theta = 1$. For a given level of labor input (the non-produced factor), Romer's equation (3.2) shows that returns to scale in the production of new knowledge using the two produced factors K and A are measured by $\beta + \theta$. Thus, our conclusion in the general R&D model is parallel to our discussion above when there was no capital: with a steadily increasing labor force, the model can converge to a steady state with constant growth only if there are diminishing returns to the produced factors. With constant or increasing returns to the produced factors, the growth rate accelerates indefinitely.

Returns to scale and endogenous growth

We have stressed several times in this chapter the importance of returns to scale in determining the properties of the model. Specifically, we have said that the long-run properties of growth models are determined by whether there are decreasing, constant, or increasing returns to scale *in the produced inputs*.

Because this issue has had a profound impact on modern growth theory, it is worth digressing to consider it in more detail. First of all, we need to be clear about what we mean by a "produced input" or "produced factor." A better term might be "endogenous input" because we consider an input to be produced if it is created endogenously within the model through the use of other factors of production. Since pure labor is exogenous in all of the growth models we have studied, it is not considered a produced factor.

In the simple Solow model of Romer's Chapter 1, advances in technology come from outside—there is no way to reallocate resources to get faster technological change—so the A term is not a produced input. However, in the R&D model of this chapter, A is produced directly by labor and capital (and A itself). Adding more resources to the R&D production function leads to more rapid accumulation of knowledge capital A. Thus, the evolution of A is endogenous and it is a produced input in this model.

Returns to scale in the produced inputs are determined by what happens to output if we multiply *only the produced inputs* by a positive constant λ. If output goes up by less than a factor of λ, then we have decreasing returns in produced inputs. We have constant returns to produced inputs is output goes up exactly by a factor of λ, and increasing returns if it increases by more than that.

"Neoclassical" growth models such as the Solow and Ramsey models have decreasing returns to produced inputs, although their production functions usually have constant returns to *all* inputs. Modern growth models have emphasized the case of constant returns to *produced* inputs (implying increasing returns to scale in all inputs), which leads to so-called ***endogenous growth***. We saw in the R&D model that changes in the economy's choice parameters, such as the saving rate and the shares of inputs devoted to R&D, lead to permanent changes in the steady-state growth rates in constant-returns models. These parameters have "growth effects" on output in endogenous growth models (*i.e.*, they change the steady-state growth rate), but only "level effects" in neoclassical models (where they affect the level, but not the slope, of the steady-state growth path).

As noted in the introduction to this chapter, economists have found endogenous growth models appealing for several reasons. First, they often lack the strong—and arguably counterfactual—convergence implications of neoclassical models. Second, many economists believe that such fundamental economic parameters as the saving rate actually have growth effects rather than just level effects on real output.

The debate over neoclassical vs. endogenous growth models has spawned a voluminous empirical literature. We shall examine a sample of this literature in Chapter 6.

Scale effects in the R&D model

One characteristic of the R&D model that may seem unrealistic at first glance is the presence of ***scale effects***. Notice in Romer's equation (3.21) that the *growth rate* of knowledge depends positively on the *level* of the population. That means that economies with large populations should grow faster than smaller ones. This result may seem surprising, but it is a direct result of the nonrival nature of knowledge in the model.

Intuitively, the more people there are in the economy, the more people can work on R&D. That will lead to the creation of more knowledge. Because knowledge is nonrival, everyone can use this knowledge to increase productivity—as discussed above, it is total knowledge that matters, not per-capita knowledge. The larger is the population, the more scientists are producing knowledge (for everyone to use) and the faster is economic growth.

International diffusion of knowledge is an important issue related to the possibility of scale effects. The relevant boundaries for the "economy" under consideration in the R&D model are the boundaries at which new knowledge stops being used. If all knowledge generated anywhere in the world is immediately used in production everywhere, then these scale effects occur on a global scale: growth in knowledge depends on the world's population. If some parts of the world economy operate in (knowledge) isolation, then knowledge in these enclaves would grow at a slower rate that is proportional to their own populations.

Based on this argument, it seems to be in the interest of every country to be integrated into a world knowledge network where knowledge moves freely. However, there are many reasons why knowledge might not transfer effectively on a worldwide scale. In practice, the use of knowledge requires substantial human capital in the using country to understand and implement the advances that have occurred. Some countries may lack the local population of engineers to apply new knowledge.[11] It is also likely that particular pieces of knowledge are more useful in some economies than in others. For example, new hybrids of crops designed for temperate regions may not help agricultural productivity in tropical areas. Advances in robotics may be irrelevant to a labor-intensive economy where robots are not used because labor is cheap relative to capital.

An interesting paper by Michael Kremer examined the plausibility of scale effects, as discussed by Romer in Section 3.7. Looking at an outrageously long time span (and correspondingly imprecise data), Kremer (1993) does indeed find that growth has been larger during periods and in places where population has been larger. Historically isolated enclaves such as Australia and Tasmania grew more slowly than large contiguous landmasses with large populations. Moreover, the overall growth rate of the economy (proxied by growth in population) seems to have accelerated over the epochs of human history as the level of population has gotten larger.

[11] This is related to the concept of "social capability" discussed by Moses Abramovitz (1986) in a paper that we shall read in the next part of the course.

D. Understanding Romer's Chapter 3, Part B

The specification of the human-capital model

The analysis of the human-capital model differs somewhat from that of the R&D model above. Because the dynamics are complex, Romer focuses on the steady state, looking at an equilibrium in which the amount of education per person is exogenous and constant. This simplifies the analysis because we can focus only on one of the two state variables: physical capital. Because K is the only state variable, the analysis turns out to be a direct extension of the Solow model.

However, this treatment ignores some important questions that are treated more carefully in the research literature. In particular, treating the level of education as an exogenous variable makes exogenous one of the central decisions of the model. Romer's simplification is analogous to the Solow model's assumption of a constant saving rate, which makes the accumulation of physical capital exogenous rather than responsive to economic incentives. A more complete specification of the model would allow individuals to decide how much human capital to accumulate based on the rate of return to education, just as agents in the Ramsey model decide on their saving (accumulation of physical capital) based on the return to capital and their desire for smooth consumption and for consumption now rather than later.

Romer begins the exposition with the production function described by equation (3.45). Note that labor input seems to be missing from the production function here. This apparent anomaly is resolved by equation (3.48), which expresses the amount of human capital $H(t)$ as the product of the number of workers $L(t)$ and a productivity factor $G(E)$ that is related to the amount of education the representative worker has received. Romer then makes the simplifying assumption that each additional year of education adds the same proportional amount to a worker's productivity, making productivity an exponential function of education as shown by equation (3.50).

Analysis with the human-capital model

To begin the analysis, we must derive the intensive form of the production function (3.45). Since $H(t)$ has taken the place of $L(t)$ in the production function, it makes sense to redefine y and k as

$$y(t) \equiv \frac{Y(t)}{A(t)H(t)} = \frac{Y(t)}{A(t)G(E)L(t)} \tag{4}$$

and

$$k(t) \equiv \frac{K(t)}{A(t)H(t)} = \frac{K(t)}{A(t)G(E)L(t)}. \tag{5}$$

With these definitions, we can write the intensive form of the Cobb-Douglas production function as

$$y(t) = k(t)^{\alpha}. \tag{6}$$

If the level of education E is constant, as Romer assumes and as it would be in a steady state; A and L are assumed to grow at constant rates as given by his equations (3.47) and (3.49); thus the denominator of (4) and (5) grows at the constant rate $n + g$, just as in the Solow model. It follows that the analysis of the intensive-form model is identical to that of the Solow model from Chapter 1. We can move immediately to write the unique, steady-state level of k as

$$k^* = \left(\frac{s}{n + g + \delta} \right)^{\frac{1}{1-\alpha}}, \tag{7}$$

as shown by Romer on page 135. From (6), the corresponding steady-state level of y is just

$$y^* = \left(\frac{s}{n + g + \delta} \right)^{\frac{\alpha}{1-\alpha}}. \tag{8}$$

In the steady state, y is constant at the level shown in equation (8), so its numerator and denominator must be growing at the same rate. With $G(E)$ constant, Y must be growing at rate $n + g$ corresponding to growth in A and L, and Y/L must grow at rate g in the steady state. Unsurprisingly, the steady-state properties of the model are entirely Solovian. The presence of $G(E)$ simply scales the level of output per worker (for given y^*) by a constant amount that depends on the equilibrium amount of education per worker.

Output per person vs. output per worker

In our analysis of the human-capital model so far, we have treated education as though it were free. People receive an amount E without paying for it by forgoing consumption of goods, labor effort, or investment in physical capital. Not surprisingly, our results suggest that more education is always better—why should we stop increasing a variable that gives us higher income and costs us nothing? But of course education is not really free. It consumes resources (mainly teacher and student time) that could otherwise be used to produce consumable output.

One way of incorporating these costs of human capital would be to think of some of our real output as being used up in education rather than being available for consumption and physical-capital investment. We can think of this approach as emphasizing the tradeoff between the use of labor and capital resources in human-capital accumulation (teachers and schools) vs. use in general production.

While this approach improves on the zero-cost assumption, it is somewhat unrealistic because the greatest cost of education is the enormous number of *student hours* that are diverted from production into human-capital accumulation. Even at Reed, which is justifiably proud of its low student/faculty ratio, there are more than ten times as many potential workers in the student body as on the faculty. Adjusting for non-teaching staff only reduces this ratio to about three.

Romer incorporates the cost of education in his model by recognizing that time spent as a student is time that you are consuming but not producing. Therefore, we must distinguish carefully between output per worker (which is the Y/L value we discussed above), and output per (adult) person, which he denotes by Y/N. The adult population N is larger than the labor force L by the number of students.

We can tell an intuitive story about why the effect of an increase in education on Y/N is going to be more complicated than the effect on Y/L. We assume that more education makes workers more productive, so increased education must lead to higher output *per worker* in the steady state. However, more education also means that a smaller share of the population is working at any given time. Since students do not produce anything, this means that the higher output of each worker is, at least partially, offset by the smaller number of workers. Romer analyzes this tradeoff beginning on page 136.

We are interested in looking at the behavior of Y/N and we know something about the behavior of $Y/L = y$, so it makes sense to start by noting that

$$\frac{Y}{N} = \frac{L}{N} \cdot \frac{Y}{L} = \frac{L}{N} y.$$

Thus, we need an expression for L/N, the share of the population that is working. This can be a little tricky if people live infinitely long, as in the Ramsey model, so Romer adopts a somewhat more realistic assumption: that everyone lives T years with the first E years devoted to education and the last $T - E$ years to working.

It might seem like we could then simply write down the ratio of working to total population as $(T - E)/T$, since that is the share of each person's life that he or she works. However, we can only do this if the population is not growing ($n = 0$). In a growing population, the young cohorts that are in education will be larger than the corresponding cohorts that are working, so the share of the population working will be somewhat smaller.

In order to calculate L/N for a growing population, we need to look explicitly at cohort size. Romer denotes the flow of people born at time t by $B(t)$. If the population is to grow at rate n with a fixed life span, then the flow of births must grow at rate n as well. Using our standard formula for continuous-time growth,

$$B(t) = B(0)e^{nt}. \tag{9}$$

We can calculate the population at time t by adding up the sizes of all cohorts born between $t - T$ and t. We are working in continuous time, so this is an integral rather than a traditional summation. Using integration to add up the flow of births from t back to $t - T$ gives

$$N(t) = \int_{\tau=0}^{T} B(t-\tau)d\tau. \tag{10}$$

This is the first part of Romer's equation (3.52). The $B(t - \tau)$ inside the integral is the flow of births that happened τ periods before time t. The population at time t includes those born between zero and T years before t, so integrating from $\tau = 0$ to $\tau = T$ adds up the cohorts that are still alive at t.

Applying equation (9) to period $t - \tau$ gives $B(t - \tau) = B(0)e^{n(t - \tau)} = B(t)e^{-n\tau}$. Romer makes this substitution to get the second line of his equation (3.52). To get the final line, he uses the rules of integrals to evaluate the integral expression. Because we have not stressed the rules of integration, a more detailed explanation is appropriate here.

Recall that integration involves "anti-differentiation," finding the function whose derivative equals the integrand. In this case, the integrand is $B(t)e^{-n\tau}$ and we are integrating with respect to τ. We can simplify the integral by noticing that $B(t)$ does not depend at all on τ, so it can be treated as a constant and brought outside the integral sign:

$$\int B(t)e^{-n\tau}d\tau = B(t)\int e^{-n\tau}d\tau. \tag{11}$$

Notice that I have temporarily suppressed the limits of integration in equation (11) and treated it as an "indefinite integral." We shall consider the limits of integration in a moment.

Remember that the derivative of the exponential function was especially simple:

$$\frac{d(e^{ax})}{dx} = ae^{ax}.$$

The anti-derivative is likewise simple,

$$\int e^{ax}dx = \frac{1}{a}e^{ax},$$

or, in this case,

$$\int e^{-n\tau}d\tau = -\frac{1}{n}e^{-n\tau}. \tag{12}$$

Equation (12) gives us the indefinite integral of the function. To calculate the definite integral over the range $\tau = 0$ to $\tau = T$ we subtract the value of the right-hand side of (12) at $\tau = 0$ from the value at $\tau = T$. Thus,

$$\int_{\tau=0}^{T} e^{-n\tau}d\tau = \left(-\frac{1}{n}e^{-nT}\right) - \left(-\frac{1}{n}e^{-n\cdot 0}\right) = \frac{1}{n}\left(e^{0} - e^{-nT}\right) = \frac{1-e^{-nT}}{n}. \tag{13}$$

Multiplying (13) by the $B(t)$ term that we took outside the integral in equation (11) yields the final expression that Romer arrives at in his equation (3.52).

The analysis of equation (3.53) to get the size of the labor force L(t) is exactly analogous. The people in the labor force at time t are the members of the population more than E years old, since people spend their first E years in education. Therefore, people born between $t - E$ and t are in school and people born between $t - T$ and $t - E$ are working. Equation (3.53) differs from equation (3.52) only in the lower limit of integration: τ ranges from E to T rather than from 0 to T.

Having derived the ratio of L to N in equation (3.54), Romer then proceeds to show in a straightforward way how output per person behaves. The principal conclusion was noted above. Increases in education have ambiguous effects on output per person. They increase output per worker but decrease ratio of workers to persons.

The specification of Romer's predation model

One of the interesting attempts to explain cross-country differences in incomes and growth has been the development of a class of models incorporating non-productive uses of resources. These models were inspired by the observation that low-income economies not only usually have less physical and human capital resources, but they often do not use the resources that they have productively. The experience of post-Communist Russia, in particular, has demonstrated the devastating economic effects that can result when talented individuals find it advantageous to use their resources in non-productive ways.

In Romer's version of the model, individuals choose whether to apply their labor to production or to "predation," which involves diverting to themselves output produced by others. A more common economic term for predatory behavior is ***rent-seeking***. Rent seekers try to make money by various schemes of redistribution (including both illegal methods like theft and such legal activities as lobbying and collecting government transfers) rather than by producing and selling goods and services. The presence of predation in the model leads to a third use of resources: protecting producers from predators, including potentially both private expenditures on protection (walls, locks, guards, and alarm systems) and social institutions such as police and courts.

As usual, we seek the simplest possible model in which we can see the particular effects of interest. There is no capital in the model, so it really is not a growth model *per se*. Rather it aims to explain cross-country differences in output between countries at a point in time (given their capital stocks), based on differences in predatory behavior. On page 160, Romer discusses some of the reinforcing effects that enter the model when capital is included.

One unfortunate feature of Romer's model in section 3.11 is that he reuses f, L, and R to represent entirely different variables than those they represented in earlier models. In this section, f is the fraction of a producer's resources that is spent on protection, L is the fraction of the producer's output that is lost to rent-seeking or predation, and R is the fraction of the population that chooses to be predators.

We choose the simplest possible production function: output equals labor input. This makes things easy by allowing us to talk interchangeably about output or labor resources, since both are measured in equivalent units. On page 155, Romer presents some assumptions about how producers' losses are affected by the number of predators and the amount of resources spent on protection. These are represented by the partial derivatives of the L function. All of these assumptions are consistent with one's common sense about how predation and protection should work.

Agents' decisions in the predation model

Each person in the model must decide whether to be a producer or a predator. This decision will depend on the returns that he or she can earn from each activity, so we must begin by deriving an expression for the amount earned in each role.

We begin by looking at the behavior and earnings of producers. An agent who has decided to be a producer faces a second decision: how much of his or her time to employ protecting output rather than producing it. The more time that is spent in protection, the larger is the share of output that is retained, but the smaller is the overall amount produced. In other words, increasing protection activity yields the producer a larger fraction of a smaller pie.

If there were no predation, then a producer with one unit of labor input would get one unit of output with no need for protection. (Recall the simple one-to-one production function we have assumed.) With predation and protection, only a fraction $(1 - f)$ of the producer's time is spent on production and only the share $(1 - L(f, R))$ of each unit of output produced is retained after predators capture

$L(f, R)$. Romer's equation (3.62) shows how the amount of output the producer keeps is related to f and R, the amount of his or her time spent on protection and the share of the population that chooses predation. Producers will choose the level of f that maximizes equation (3.62), so we take the derivative of (3.62) with respect to f and set it equal to zero. The result is equation (3.64), which gives f as an ***implicit function*** of R.[12]

Common sense suggests that an increase in the number of predators will lead to an increase in producers' protection efforts. We can show that this is true by calculating df/dR, the effect of a change in R on f. If we could solve (3.64) for an explicit function with f appearing only on the left-hand side, this derivative could be calculated in the usual way. However, with f as an implicit function, we can still find df/dR by using the technique of ***implicit differentiation***.

To perform implicit differentiation, we take the derivative of both sides of either (3.63) with respect to R. We can choose to work with either (3.63) or (3.64) because they are equivalent, but (3.63) is easier to differentiate because it does not involve quotients. If f and R were independent, then we could take a partial derivative with respect to R and treat f as a constant. However, because f depends on R, each time we take the derivative of f with respect to R we get df/dR.

To see how implicit differentiation works, let's call the left-hand side of Romer's equation (3.62) $Z(f, R)$:

$$Z(f,R) \equiv -[1-L(f,R)]-(1-f)L_f(f,R).$$

The rule for implicit differentiation when f is a function of R is that

$$\frac{dZ(f,R)}{dR} = \frac{\partial Z(f,R)}{\partial f}\frac{df}{dR} + \frac{\partial Z(f,R)}{\partial R}. \tag{14}$$

The first term in (14) applies the chain rule, recognizing that Z is a function of f and that f is a function of R. Note that there are very similar looking expressions on the left and right sides of equation (14) that are actually quite different. The ***total derivative*** on the left side is the total effect of R on Z taking all channels into account. The second term on the right side is a *partial derivative* (with the ∂ symbol rather than d), which is the effect of R on Z holding the other argument of Z (in this case, f) constant. In English, equation (14) says that the total effect of R on Z is the partial (indirect) effect of f on Z times the effect of R on f, plus the partial (direct) effect of R on Z.

[12] An implicit function is one where we cannot solve to isolate the dependent variable alone on the left-hand side of the equation. Because f is "inside" the L function, whose form is unspecified, there is no way to get f alone in equation (3.63).

To evaluate the effect of R on f in equation (3.63), we use the rules of differentiation to evaluate the two partial derivatives in (14), using the function on the left side of (3.63) as the Z function. This leads to Romer's equation (3.65). Solving (3.65) for df/dR yields equation (3.66), which is the expression we seek.

To establish that an increase in the number of predators raises protection effort, we must determine the sign of the right side of (3.66). The assumptions that Romer makes on page 155 are almost sufficient to do so.[13] In particular, $L_R > 0$, $(1 - f)$ is positive, and $L_{fR} < 0$, so the numerator is a positive minus a negative, which is surely positive. Since $L_{ff} > 0$ and $L_f < 0$, the denominator is also a positive number minus a negative one. With numerator and denominator both positive, the quotient is also positive and $df/dR > 0$. Thus we can write the relationship between f and R as a new function $f = f(R)$, with $f' > 0$. The amount of time devoted to protection depends positively on the number of rent-seeking predators.

Having determined the nature of the producer's optimal protection behavior, we next turn to the decision of whether an individual should be a producer or a predator. As usual in economics, we assume that individuals will choose whichever alternative gives them the higher return.[14] If, given the current number of predators and producers, the return on production was higher (lower) than that on predation, then some agents would shift from predation to production (production to predation). Thus, the only possible (interior) equilibrium is where the returns on the two activities are equal.[15]

Romer's (3.67) equates the returns to the two activities. The figures on pages 157 and 158 show some possible patterns of equilibria, including the interesting case of multiple equilibria in Figure 3.11.

In order for the predation model to be useful in explaining cross-country income differentials, we must be able to explain how differences in the amount of resources devoted to production, $(1 - f)\,(1 - R)$, depend on institutional characteristics of an economy. Romer does this on pages 158 and 159 by considering the effects of differences in the ease of predation. He puts this in terms of the probability of a predator being caught and having his or her earnings taken away. This would be appropriate when the predatory rent-seeking in question is illegal. However, we could equally

[13] If we strengthen the assumptions to rule out zero partial derivatives, as we shall do, then df/dR is surely positive. With the weak inequality assumptions a zero value is possible.

[14] That means that people in this model have no "conscience"—there is no disutility attached to being a predator.

[15] A corner solution would occur if the returns to production were higher (lower) than those to predation even if there were no predators (producers) in the economy. Then no one would choose to be a predator (producer). There can be no corner solution in which all are predators because there would be nothing to steal.

well apply this model to legal forms of predation by simply varying the ease with which rent-seekers can claim others' resources. For example, a very generous system of unemployment compensation or welfare could be thought of as an easy-predation regime, as could a political system in which advantages such as monopoly franchises can easily be obtained by lobbying or bribes.

The discussion on pages 158 through 160 demonstrates that a decrease in the returns to predation leads to a multiplier effect on output. An initial decrease in the number of predators raises the returns to both production and predation, but (the way Figure 3.12 is drawn) increases production returns by more, so more people will shift from predation to production. Moreover, when fewer resources are devoted to predation, producers will also be able to spend more time producing and less time protecting their output from predators.

An additional, very important effect of predation is described on page 160. Individuals are extremely reluctant to invest in capital in an economy where predation may expropriate it at some future date. This is the reason for the vast "capital flight" from countries in which property rights are insecure. Making prospective investors, whether foreigners or domestic residents who are currently sending their funds abroad, feel more secure about property rights in their investments is likely to increase capital accumulation, leading to higher incomes.

E. Suggestions for Further Reading

General texts on modern growth theory

Barro, Robert J., and Xavier Sala-i-Martin, *Economic Growth,* 2nd ed., Cambridge, Mass.: MIT Press, 2004, Chapters 4 through 7.

Aghion, Philippe, and Peter Howitt, *Endogenous Growth Theory* Cambridge, Mass.: MIT Press, 1998.

Acemoglu, Daron, *Introduction to Modern Economic Growth*, Princeton, N.J.: Princeton University Press, 2009. (An encyclopedic and highly mathematical new textbook on growth theory.)

Jones, Charles I., *Introduction to Economic Growth* (New York: W.W. Norton, 1998), Chapters 4 through 8. (A much simpler mathematical treatment.)

Selected seminal papers in modern growth theory

Romer, Paul M., "Increasing Returns and Long-Run Growth," *Journal of Political Economy* 94(5), October 1986, 1002–1037. (The paper that is generally regarded as having started it all.)

Lucas, Robert E., Jr., "On the Mechanics of Development Planning," *Journal of Monetary Economics* 22(1), July 1988, 3–42. (Lucas's Nobel address focused on a two-sector model with human capital.)

Romer, Paul M., "Endogenous Technical Change," *Journal of Political Economy* 98(5), October 1990, Part II, S71–S102. (The Paul Romer model closest to the R&D model in David Romer's Chapter 3.)

Aghion, Philippe, and Peter Howitt, "A Model of Growth Through Creative Destruction," *Econometrica* 60(2), March 1992, 323–351. (A seminal paper in the neo-Schumpeterian strain of endogenous growth models.)

Grossman, Gene M., and Elhanan Helpman, *Innovation and Growth in the Global Economy* (Cambridge, Mass.: MIT Press, 1991). (A book elaborating on a variety of endogenous growth models in an international context.)

Rebelo, Sergio, "Long-Run Policy Analysis and Long-Run Growth," *Journal of Political Economy* 99(3), June 1991, 500–521. (One of the most often cited *AK* models.)

F. Works Cited in Text

Abramovitz, Moses. 1986. Catching Up, Forging Ahead, and Falling Behind. *Journal of Economic History* 46 (2):385-406.

Acemoglu, Daron. 2009. *Introduction to Modern Economic Growth*. Princeton, N.J.: Princeton University Press.

Barro, Robert J., and Xavier Sala-i-Martin. 2004. *Economic Growth*. 2nd ed. Cambridge, Mass.: MIT Press.

Klenow, Peter J. 1998. Ideas versus Rival Human Capital: Industry Evidence on Growth Models. *Journal of Monetary Economics* 42 (1):3-23.

Kremer, Michael. 1993. Population Growth and Technological Change: One Million B.C. to 1990. *Quarterly Journal of Economics* 108 (3):681-716.

Landes, David S. 1983. *Revolution in Time: Clocks and the Making of the Modern World*. Cambridge, Mass.: Harvard University Press.