Chapter 16

GROWTH AND IDEAS

CHARLES I. JONES

Department of Economics, University of California, Berkeley
and
NBER

Contents

Handbook of Economic Growth, Volume 1B. Edited by Philippe Aghion and Steven N. Durlauf

DOI: 10.1016/S1574-0684(05)01016-6

Abstract

Ideas are different from nearly all other economic goods in that they are nonrivalrous. This nonrivalry implies that production possibilities are likely to be characterized by increasing returns to scale, an insight that has profound implications for economic growth. The purpose of this chapter is to explore these implications.

Keywords

economic growth, ideas, scale effects, survey

JEL classification: O40, E10

1. Introduction

People in countries like the United States are richer by a factor of about 10 or 20 than people a century or two ago. Whereas U.S. per capita income today is $33,000, conventional estimates put it at $1800 in 1850. Yet even this difference likely understates the enormous increase in standards of living over this period. Consider the quality of life of the typical American in the year 1850. Life expectancy at birth was a scant 40 years, just over half of what it is today. Refrigeration, electric lights, telephones, antibiotics, automobiles, skyscrapers, and air conditioning did not exist, much less the more sophisticated technologies that impact our lives daily in the 21st century.[1]

Perhaps the central question of the literature on economic growth is "Why is there growth at all?" What caused the enormous increase in standards of living during the last two centuries? And why were living standards nearly stagnant for the thousands and thousands of years that preceded this recent era of explosive growth?

The models developed as part of the renaissance of research on economic growth in the last two decades attempt to answer these questions. While other chapters discuss alternative explanations, this chapter will explore theories in which the economics of ideas takes center stage. The discoveries of electricity, the incandescent lightbulb, the internal combustion engine, the airplane, penicillin, the transistor, the integrated circuit, just-in-time inventory methods, Wal–Mart's business model, and the polymerase chain reaction for replicating strands of DNA all represent new ideas that have been, in part, responsible for economic growth over the last two centuries.

The insights that arise when ideas are placed at the center of a theory of economic growth can be summarized in the following Idea Diagram:

$$\text{Ideas} \Rightarrow \text{Nonrivalry} \Rightarrow \text{IRS} \Rightarrow \text{Problems with CE.}$$

To understand this diagram, first consider what we mean by "ideas". Romer (1993) divides goods into two categories: ideas and objects. Ideas can be thought of as instructions or recipes, things that can be codified in a bitstring as a sequence of ones and zeros. Objects are all the rivalrous goods we are familiar with: capital, labor, output, computers, automobiles, and most fundamentally the elemental atoms that make up these goods. At some level, ideas are instructions for arranging the atoms and for using the arrangements to produce utility. For thousands of years, silicon dioxide provided utility mainly as sand on the beach, but now it delivers utility through the myriad of goods that depend on computer chips. Viewed this way, economic growth can be sustained even in the presence of a finite collection of raw materials as we discover better ways to arrange atoms and better ways to use the arrangements. One then naturally wonders about possible limits to the ways in which these atoms can be arranged, but

[1] Ideally, the calculations of GDP should take the changing basket of goods and changes in life expectancy into account, but the standard price indices used to construct these comparisons are inadequate. See, for example, DeLong (2000) and Nordhaus (2003).

the combinatorial calculations of Romer (1993) and Weitzman (1998) quickly put such concerns to rest. Consider, for example, the number of unique ways of ordering twenty objects (these could be steps in assembling a computer chip or ingredients in a chemical formula). The answer is 20!, which is on the order of 10^{18}. To put this number in perspective, if we tried one different combination every second since the universe began, we would have exhausted less than twenty percent of the possibilities.[2]

The first arrow in the Idea Diagram links ideas with the concept of nonrivalry. Recall from public economics that a good is nonrivalrous if one person's use of the good does not diminish another's use. Most economic goods – objects – are rivalrous: one person's use of a car, a computer, or an atom of carbon dimishes the ability of someone else to use that object. Ideas, by contrast, are nonrivalrous. As examples, consider public key cryptography and the famous introductory bars to Beethoven's Fifth Symphony. Audrey's use of a particular cryptographic method does not inhibit my simultaneous use of that method. Nor does Benji's playing of the Fifth Symphony limit my (in)ability to perform it simultaneously. For an example closer to our growth models, consider the production of computer chips. Once the design of the latest computer chip has been invented, it can be applied in one factory or two factories or ten factories. The design does not have to be reinvented every time a new computer chip gets produced – the same idea can be applied over and over again. More generally, the set of instructions for combining and using atoms can be used at any scale of production without being diminished.

The next link between nonrivalry and increasing returns to scale (IRS) is the first indication that nonrivalry has important implications for economic growth. As discussed in Romer (1990), consider a production function of the form

$$Y = F(A, X), \tag{1}$$

where Y is output, A is an index of the amount of knowledge that has been discovered, and X is a vector of the remaining inputs into production (e.g. capital and labor). Our standard justification for constant returns to scale comes from a replication argument. Suppose we'd like to double the production of computer chips. One way to do this is to replicate all of the standard inputs: we build another factory identical to the first and populate it with the same material inputs and with identical workers. Crucially, however, we do not need to double the stock of knowledge because of its nonrivalry: the existing design for computer chips can be used in the new factory by the new workers.

One might, of course, require additional copies of the blueprint, and these blueprints may be costly to produce on the copying machine down the hall. The blueprints are not ideas; the copies of the blueprints might be thought of as one of the rivalrous inputs included in the vector X. The bits of information encoded in the blueprint – the design for the computer chip – constitute the idea.

[2] Of course, one also must consider the fraction of combinations that are useful. Responding to one such combinatorial calculation, George Akerlof is said to have wondered, "Yes, but how many of them are like chicken ice cream?".

Mathematically, we can summarize these insights in the following two equations. For some number $\lambda > 1$,

$$F(A, \lambda X) = \lambda Y, \tag{2}$$

and as long as more knowledge is useful,

$$F(\lambda A, \lambda X) > \lambda Y. \tag{3}$$

That is, there are constant returns to scale to the standard rivalrous inputs X and, therefore, increasing returns to scale to these inputs and A taken together. If we double the number of factories, workers, and materials *and* double the stock of knowledge, then we will more than double the production of computer chips. Including ideas as an input into production naturally leads one to models in which increasing returns to scale plays an important role. Notice that a "standard" production function in macroeconomics of the form $Y = K^\alpha (AL)^{1-\alpha}$ builds in this property.

Introducing human capital into this framework adds an important wrinkle but does not change the basic insight. Suppose that the design for a computer chip must be learned by a team of scientists overseeing production before it can be used, thus translating the idea into human capital. To double production, one can double the number of factories, workers, and scientists. If one incorporates a better-designed computer chip as well, production more than doubles. Notice that the human capital is rivalrous: a scientist can work on my project or your project, but not on both at the same time. In contrast, the idea is nonrivalrous: two scientists can both implement a new design for a computer chip simultaneously.

Confusion can arise in thinking about human capital if one is not careful. For example, consider a production function that is constant returns in physical and human capital, the two rivalrous inputs: $Y = K^\alpha H^{1-\alpha}$. Now, suppose that $H = hL$, where h is human capital per person. Then, this production function is $Y = K^\alpha (hL)^{1-\alpha}$. There were constant returns to K and H in our first specification, but one is tempted conclude that there are increasing returns to K, L, and h together in the rewritten form. Which is it? Does the introduction of human capital involve increasing returns, just like the consideration of ideas?

The answer is no. To see why, consider a different example, this time omitting human capital altogether. Suppose $Y = K^\alpha L^{1-\alpha}$. This is perhaps our most familiar Cobb–Douglas production function and it exhibits constant returns to scale in K and L. Now, rewrite this production function as $Y = k^\alpha L$, where $k \equiv K/L$ is physical capital per person. Would we characterize this production function as possessing increasing returns? Of course not! Obviously a simple change of variables cannot change the underlying convexity of a production function.

This example suggests the following principle. In considering the degree of homogeneity of a production function, one must focus on the function that involves total quantities, so that nothing is "per worker". Intuitively, this makes sense: if one is determining returns to scale, the presence of "per worker" variables will of course lead to confusion. The application of this principle correctly identifies the production function based on human capital $Y = K^{\alpha} H^{1-\alpha}$ as possessing constant returns. Introducing ideas into the production function leads to increasing returns because of nonrivalry.

Finally, the last link in our diagram connects increasing returns to scale to "Problems with CE", by which we mean problems with the standard decentralization of the optimal allocation of resources using a perfectly competitive equilibrium. A central requirement of a competitive equilibrium is that factors get paid their marginal products. But with increasing returns to scale, as is well known, this is not possible. Continuing with the production function in Equation (1), the property of constant returns in X guarantees that[3]

$$F_X X = Y. \tag{4}$$

That is, paying each rivalrous factor its marginal product exhausts output, so that nothing would be left over to compensate the idea inputs

$$F_X X + F_A A > Y. \tag{5}$$

If the stock of knowledge is also paid its marginal product, then the firm would make negative profits. This means that the standard competitive equilibrium will run into problems in a model that includes ideas.

These two implications of incorporating ideas into our growth models – increasing returns and the failure of perfect competition to deliver optimal allocations – are the basis for many of the insights and results that follow in the remainder of this chapter. This chain of reasoning provides the key foundation for idea-based growth theory.

The purpose of this chapter is to outline the contribution of idea-based growth models to our understanding of economic growth. The next section begins by providing a brief overview of the intellectual history of idea-based growth theory, paying special attention to developments that preceded the advent of new growth theory in the mid-1980s. Section 3 presents the simplest possible model of growth and ideas in order to illustrate how these theories explain long-run growth. Section 4 turns to a richer model. This framework is used to compare the allocation of resources in equilibrium with the optimal allocation. The richer model also serves as the basis for several applications that follow in Sections 5 and 6. Section 5 provides a discussion of the scale effects that naturally emerge in models in which ideas play an important role and reviews a number of related contributions. Section 6 summarizes what we have learned from growth accounting in idea-based growth models, considers a somewhat controversial criticism of endogenous growth models called the "linearity critique", and briefly summarizes some of the additional literature on growth and ideas. Finally, Section 7 of this chapter concludes by discussing several of the most important open questions related to growth and ideas.

[3] Since X is a vector, the notation in this equation should be interpreted as the dot product between the vector of derivatives and the vector of inputs.

It is worth mentioning briefly as well what this chapter omits. The most significant omission is a careful presentation of the Schumpeterian growth models of Aghion and Howitt (1992) and Grossman and Helpman (1991) and the very interesting directions in which these models have been pushed. This omission, however, is remedied in another chapter of this Handbook by Aghion and Howitt. Probably the next most important omission is a serious discussion of the empirical work in what is known as the productivity literature on the links between R&D, growth, and social rates of return. An excellent overview of this literature can be found in Griliches (1998).

2. Intellectual history of this idea

The fundamental insight conveyed by the Idea Diagram is an idea itself. And like many ideas, it is one that has been discovered, at least in part, several times in the past, at times being appreciated as a deep insight and at times being forgotten. A brief intellectual history of this idea follows, in part because it is useful to document this history but also in part because it helps to illuminate the idea itself.

William Petty, an early expert on the economics of taxation, identified in 1682 one of the key benefits of a larger population:

> 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 400 persons. [Quoted by Simon (1998), p. 372.]

More than a century later, Thomas Jefferson came closer to characterizing the nonrivalrous nature of an idea:[4]

> Its peculiar character . . . is that no one possesses the less, because every other possesses the whole of it. He who receives an idea from me, receives instruction himself without lessening mine; as he who lights his taper at mine, receives light without darkening me. That ideas should freely spread from one to another over the globe, for the moral and mutual instruction of man, and improvement of his condition, seems to have been peculiarly and benevolently designed by nature, when she made them, like fire, expansible over all space, without lessening their density at any point . . . [Letter from Thomas Jefferson to Isaac McPherson, August 13, 1813, collected in Lipscomb and Bergh (1905), pp. 333–335.]

But it was not until the 1960s that economists systematically explored the economics of ideas. Kuznets (1960) intuits a link between population, ideas, and economic growth, and Boserup (1965) emphasizes how population pressure can lead to the adoption of

[4] David (1993) cites this passage in emphasizing that ideas are "infinitely expansible", a phrase picked up by Quah (1996).

new technologies. Arrow (1962b) and Shell (1966) clearly recognize the failure of models of perfect competition to deliver optimal resource allocation in the presence of ideas. Phelps (1966) and Nordhaus (1969) present explicit models in which the nonrivalry of knowledge leads to increasing returns and derive the result, discussed in detail below, that long-run growth in per capita income is driven by population growth.[5] Still, neither of these papers knows quite how seriously to take this prediction, with Nordhaus calling it a "peculiar result" (p. 23). Within two years, however, Phelps (1968) is convinced:

> One can hardly imagine, I think, how poor we would be today were it not for the rapid population growth of the past to which we owe the enormous number of technological advances enjoyed today... If I could re-do the history of the world, halving population size each year from the beginning of time on some random basis, I would not do it for fear of losing Mozart in the process. [Pp. 511–512.]

This implication then becomes central to the popular writings of Julian Simon in the debates over the merits and drawbacks of population growth, as in Simon (1986, 1998).

The formal literature on idea-based growth falters considerably in the 1970s and early 1980s. Much of the work that is carried out involves applications of the basic Solow (1956) model and the growth accounting calculations that subsequently followed. By the mid-1980s, many of the insights gleaned during the 1960s were no longer being taught in graduate programs. In part, this period of neglect seems to have stemmed from a lack of adequate techniques for modeling the departures from perfect competition that are implied by the economics of ideas [e.g. see Romer (1994b)]. This theoretical gap gets filled through the work on imperfect competition by Spence (1976) and Dixit and Stiglitz (1977).

Idea-based growth models are thrust to center stage in the profession with the publication of a series of papers by Romer (1986, 1987, 1990). These papers – most especially the last one – lay out with startling clarity the link between economic growth and ideas.[6] Shortly thereafter, the models of Aghion and Howitt (1992) and Grossman and Helpman (1991) introduce the Schumpeterian notions of creative destruction and business stealing, pushing idea-based growth theory further.[7]

3. A simple idea-based growth model

3.1. The model

It is useful to begin with the simplest possible idea-based growth model in order to see clearly how the key ingredients fit together to provide an explanation of long-run

[5] The learning-by-doing models of Arrow (1962a) and Sheshinski (1967) contain a similar result.

[6] This brief review obviously ignores many fundamental contributions to growth theory in order to focus on the history of idea-based growth models. Other chapters in this Handbook will lay out the roles played by neoclassical growth models, AK models, and models of growth driven by human capital accumulation.

[7] Other important contributions around this time include Judd (1985) and Segerstrom, Anant and Dinopoulos (1990).

growth. To strip the model to its essence, we ignore physical capital and human capital; these will be introduced in the richer framework of Section 4.

Suppose that in our toy economy the only rivalrous input in production (the X variable in the Introduction) is labor. The economy contains a single consumption good that is produced according to

$$Y_t = A_t^{\sigma} L_{Yt}, \quad \sigma > 0, \tag{6}$$

where Y is the quantity of output of the good, A is the stock of knowledge or ideas, and L_Y is the amount of labor used to produce the good. Notice that there are constant returns to scale to the rivalrous inputs, here just labor, and increasing returns to labor and ideas taken together. To double the production of output, it is sufficient to double the amount of labor using the same stock of knowledge. If we also double the stock of knowledge, we would more than double output.

The other good that gets produced in this economy is knowledge itself. Just as more workers can produce more output in Equation (6), more researchers can produce more new ideas:

$$\dot{A}_t = \nu(A_t) L_{At} = \nu L_{At} A_t^{\phi}, \quad \nu > 0. \tag{7}$$

If A is the stock of knowledge, then $\dot{A}$ is the amount of new knowledge produced at time t. L_A denotes the number of researchers, and each researcher can produce $\nu(A)$ new ideas at a point in time. To simplify further, we assume that $\nu(A)$ is a power function.

Notice the similarity between Equations (6) and (7). Both equations involve constant returns to scale to the rivalrous labor input, and both allow departures from constant returns because of the nonrivalry of ideas. Ideas are simply another good in this economy that labor can produce.

If $\phi > 0$, then the number of new ideas a researcher invents over a given interval of time is an increasing function of the existing stock of knowledge. We might label this the *standing on shoulders effect*: the discovery of ideas in the past makes us more effective researchers today. Alternatively, though, one might consider the case where $\phi < 0$, i.e. where the productivity of research declines as new ideas are discovered. A useful analogy in this case is a fishing pond. If the pond is stocked with only 100 fish, then it may be increasingly difficult to catch each new fish. Similarly, perhaps the most obvious new ideas are discovered first and it gets increasingly difficult to find the next new idea.

With these production functions given, we now specify a resource constraint and a method for allocating resources. The number of workers and the number of researchers sum to the total amount of labor in the economy, L,

$$L_{Yt} + L_{At} = L_t. \tag{8}$$

The amount of labor, in turn, is assumed to be given exogenously and to grow at a constant exponential rate n,

$$L_t = L_0 e^{nt}, \quad n > 0. \tag{9}$$

Finally, the only allocative decision that needs to be made in this simple economy is how to allocate labor. We make a Solow-like assumption that a constant fraction s of the labor force works as researchers, leaving $1 - s$ to produce goods.

3.2. Solving for growth

The specification of this economy is now complete, and it is straightforward to solve for growth in per capita output, $y \equiv Y/L$. First, notice the important result that $y_t = (1 - s)A_t^{\sigma}$, i.e. per capita output is proportional to the stock of ideas (raised to some power). Because of the nonrivalry of ideas, per capita output depends on the *total* stock of ideas, not on the stock of ideas per capita.

Taking logs and time derivatives, we have the corresponding relation in growth rates

$$\frac{\dot{y}_t}{y_t} = \sigma \frac{\dot{A}_t}{A_t}. \tag{10}$$

Growth of per capita output is proportional to the growth rate of the stock of knowledge, where the factor of proportionality measures the degree of increasing returns in the goods sector.

The growth rate of the stock of ideas, in turn is given by

$$\frac{\dot{A}_t}{A_t} = \nu \frac{L_{At}}{A_t^{1-\phi}}. \tag{11}$$

Under the assumption that $\phi < 1$, it is straightforward to show that the dynamics of this economy lead to a stable balanced growth path (defined as a situation in which all variables grow at constant rates, possibly zero). For the growth rate of A to be constant in Equation (11), the numerator and denominator of the right-hand side of that equation must grow at the same rate. Letting g_x denote the growth rate of some variable x along the balanced growth path, we then have

$$g_A = \frac{n}{1 - \phi}. \tag{12}$$

The growth rate of the stock of ideas, in the long-run, is proportional to the rate of population growth, where the factor of proportionality depends on the degree of returns to scale in the production function for ideas.

Finally, this equation can be substituted into Equation (10) to get the growth rate of output per worker in steady state,

$$g_y = \sigma g_A = \frac{\sigma n}{1 - \phi}. \tag{13}$$

The growth rate of per capita output is proportional to the rate of population growth, where the factor of proportionality depends on the degree of increasing returns in the two sectors.

3.3. Discussion

Why is this the case? There are two basic elements of the toy economy that lead to the result. First, just as the total output of any good depends on the total number of workers producing the good, more researchers produce more new ideas. A larger population means more Mozarts and Newtons, and more Wright brothers, Sam Waltons, and William Shockleys. Second, the nonrivalry of knowledge means that per capita output depends on the total stock of ideas, not on ideas per person.[8] Each person in the economy benefits from the new ideas created by the Isaac Newtons and William Shockleys of the world, and this benefit is not degraded by the presence of a larger population.

Together, these steps imply that output per capita is an increasing function, in the long run, of the number of researchers in the economy, which in turn depends on the size of the population. Log-differencing this relation, the growth rate of output per capita depends on the growth rate of the number of researchers, which in turn is tied to the rate of population growth in the long run.

At some basic level, these results should not be surprising at all. Once one grants that the nonrivalry of ideas implies increasing returns to scale, it is nearly inevitable that the size of the population affects the level of per capita income. After all, that is virtually the definition of increasing returns.

In moving from this toy model to the real world, one must obviously be careful. Probably the most important qualification is that our toy model consists of a single country. Without thinking more carefully about the flows of ideas across countries in the real world, it is more accurate to compare the predictions of this toy economy to the world as a whole rather than to any single economy. Taiwan and China both benefit from ideas created throughout the world, so it is not the Taiwanese or Chinese population that is especially relevant to those countries' growth experiences.

Another qualification relates to the absence of physical and human capital from the model. At least as far as long-run growth is concerned, this absence is not particularly harmful: recall the intuition from the Solow growth model that capital accumulation is not, by itself, a source of long-run growth. Still, because of transition dynamics these factors are surely important in explaining growth over any given time period, and they will be incorporated into the model in the next section.

Finally, it is worth mentioning briefly how this result differs from the original results in the models of Romer (1990), Aghion and Howitt (1992), and Grossman and Helpman (1991). Those models essentially make the assumption that $\phi = 1$ in the production function for new ideas. That is, the growth rate of the stock of knowledge depends on the number of researchers. This change serves to strengthen the importance of increasing returns to scale in the economy, so much so that a growing number of researchers causes the growth rate of the economy to grow exponentially. We will discuss this result in more detail in later sections.

[8] Contrast this to the case in which "capital" replaces the word "ideas" in this phrase. Because capital is rivalrous, output per capita depends on capital per person.

4. A richer model and the allocation of resources

The simple model given in the previous section provides several of the key insights of idea-based growth models, but it is too simple to provide others. In particular, the final implication in the basic Idea Diagram related to the problems a competitive equilibrium has in allocating resources has not been discussed. In this section, we remedy this shortcoming and discuss explicitly several mechanisms for allocating resources in an economy in which ideas play a crucial role. In addition, we augment the simple model with the addition of physical capital, human capital, and the Dixit–Stiglitz love of variety approach that has proven to be quite useful in modeling growth.

The model presented in this section is developed in a way that has become a de facto standard in macroeconomics. First, the economic environment – the collection of production technologies, resource constraints, and utility functions – is laid out. Any method of allocating resources is constrained by the economic environment. Next, we present several different ways in which resources can be allocated in this economy and derive results for each allocation. The first allocation is the simplest: a rule-of-thumb allocation analogous to the constant saving rate assumption of Solow (1956). The second allocation is the optimal one, i.e. the allocation that maximizes utility subject to the constraints imposed by the economic environment. These first two are very natural allocations to consider. One then immediately is led to ask the question of whether a decentralized equilibrium allocation, that is one in which markets allocate resources rather than a planner, can replicate the optimal allocation. In general, the answer to this question is that it depends on the nature of the institutions that govern the equilibrium. We will solve explicitly for one of these equilibrium allocations in Section 4.4 and then discuss several alternative institutions that might be used to allocate resources in this model.

4.1. The economic environment

The economic environment for this new model consists of a set of production functions, a set of resource constraints, and preferences. These will be described in turn.

First, the basic production functions are these:

$$Y_t = \left(\int_0^{A_t} x_{it}^{\theta}\,\mathrm{d}i\right)^{\alpha/\theta} H_{Yt}^{1-\alpha}, \quad 0 < \alpha < 1,\ 0 < \theta < 1, \tag{14}$$

$$\dot{K}_t = Y_t - C_t - \delta K_t, \quad K_0 > 0,\ \delta > 0, \tag{15}$$

$$\dot{A}_t = \nu H_{At}^{\lambda} A_t^{\phi}, \quad A_0 > 0,\ \nu > 0, \lambda > 0,\ \phi < 1. \tag{16}$$

Equation (14) is the production function for the final output good. Final output Y is produced using human capital H_Y and a collection of intermediate capital goods x_i. A represents the measure of these intermediate goods that are available at any point in time. These intermediate goods enter the production function through a CES aggregator function, and the elasticity of substitution between intermediate goods is $1/(1-\theta) > 1$.

Notice that there are constant returns to scale in H_Y and these intermediate goods in producing output for a given A. However, there are increasing returns to scale once A is treated as a variable. The sense in which this is true will be made precise below.

Equation (15) is a standard accumulation equation for physical capital.

Equation (16) is the production function for new ideas. In this economy, ideas have a very precise meaning – they represent new varieties of intermediate goods that can be used in the production of final output. New ideas are produced with a Cobb–Douglas function of human capital and the existing stock of knowledge.[9] As in the simple model, the parameter ϕ measures the way in which the current stock of knowledge affects the production of new ideas. It nets out the standing on shoulders effect and the fishing out effect. The parameter λ represents the elasticity of new idea production with respect to the number of researchers. A value of $\lambda = 1$ implies that doubling the number of researchers doubles the production of new ideas at a point in time for a given stock of knowledge. On the other hand, one imagines that doubling the number of researchers might less than double the number of new ideas because of duplication, suggesting $\lambda < 1$.

Next, the resource constraints for the economy are given by

$$\int_0^{A_t} x_{it}\,\mathrm{d}i = K_t, \tag{17}$$

$$H_{At} + H_{Yt} = H_t, \tag{18}$$

$$H_t = h_t L_t, \tag{19}$$

$$h_t = \mathrm{e}^{\psi \ell_{ht}}, \quad \psi > 1, \tag{20}$$

$$L_t = (1 - \ell_{ht}) N_t, \tag{21}$$

$$N_t = N_0 \mathrm{e}^{nt}, \quad N_0 > 0,\ n > 0. \tag{22}$$

Breaking slightly from my taxonomy, Equation (17) involves a production function as well as a resource constraint. In particular, one unit of raw capital can be transformed instantaneously into one unit of any intermediate good for which a design has been discovered. Equation (17) then is the resource constraint that says that the total quantity of intermediate goods produced cannot exceed the amount of raw capital in the economy.

Equation (18) says that the amount of human capital used in the production of goods and ideas equals the total amount of human capital available in the economy. Equation (19) states the identity that this total quantity of human capital is equal to human capital per person h times the total labor force L (all labor is identical). An individual's

[9] Physical capital is not used in the production of new ideas in order to simplify the model. A useful alternative to this approach is the "lab equipment" approach suggested by Rivera-Batiz and Romer (1991) where units of the final output good are used to produce ideas, i.e. capital and labor combine in the same way to produce ideas as to produce final output. Apart from some technicalities, all of the results given below have exact analogues in a lab-equipment approach.

human capital is related by the Mincerian exponential to the amount of time spent accumulating human capital, ℓ_h, in Equation (20). We simplify the model by assuming there are no dynamics associated with human capital accumulation.[10] Equation (21) defines the labor force to be the population multiplied by the amount of time that people are not accumulating human capital, and Equation (22) describes exogenous population growth at rate n.

Finally, preferences in this economy take the usual form:[11]

$$U_t = \int_t^\infty N_s u(c_s) \mathrm{e}^{-\rho(s-t)}\,\mathrm{d}s, \qquad \rho > n, \tag{23}$$

$$c_t \equiv \frac{C_t}{N_t}, \tag{24}$$

$$u(c) = \frac{c^{1-\zeta} - 1}{1-\zeta}, \qquad \zeta > 0. \tag{25}$$

4.2. Allocating resources with a rule of thumb

Given this economic environment, we can now consider various ways in which resources may be allocated. The primary allocative decisions that need to be made are relatively few. At each point in time, we need to determine the amount of time spent gaining human capital ℓ_h, the amount of consumption c, the amount of human capital allocated to research H_A, and the split of the raw capital into the various varieties $\{x_i\}$. Once these allocative decisions have been made, the twelve equations in (14) to (25) above, combined with these four allocations pin down all of the quantities in the model.[12]

The simplest way to begin allocating resources in just about any model is with a "rule of thumb". That is, the modeler specifies some simple, exogenous rules for allocating resources. This is useful for a number of reasons. First, it forces us to be clear from

10 This approach can be justified by a simple dynamic system of the form $\dot{h} = \mu \mathrm{e}^{\psi \ell_h} - \delta h$, where human capital depreciates at rate δ. It is readily seen that in the steady state, this equation implies that h is proportional to $\mathrm{e}^{\psi \ell_h}$, as we have assumed. More generally, of course, richer equations for human capital can be imagined.

11 To keep utility finite, we require a technical condition on the parameters of the model. The appropriate condition can be determined by looking at the utility function and takes the form

$$\rho > n + \frac{\lambda}{1-\phi} \frac{\sigma}{1-\alpha} (1-\zeta) n.$$

12 The counting goes as follows. At a point in time we have the four allocation rules and the twelve equations given above. The four rules pin down the allocations ℓ_h, C, H_A, $\{x_i\}$ and then twelve equations deliver Y, K, A, H_Y, H, h, L, N, U, c and u. The careful counter will notice I have mentioned 15 objects but 16 equations. The subtlety is that we should think of the allocation rule as determining $\{x_i\}$ subject to the resource constraint in (17). (For comparison, notice that we choose H_A and the resource constraint pins down H_Y. Similarly, but loosely speaking, we choose "all but one" of the x_i and the resource constraint pins down the last one.)

the beginning about exactly what allocation decisions need to be made. Second, it reveals how key endogenous variables depend on the allocations themselves. This is nice because the subsequent results will hold along a balanced growth path even if other mechanisms are used to allocated resources.

DEFINITION 4.1. A *rule of thumb allocation* in this economy consists of the following set of equations:

$$\ell_{ht} = \bar{\ell}_h \in (0,1), \tag{26}$$

$$1 - \frac{C_t}{Y_t} = \bar{s}_K \in (0,1), \tag{27}$$

$$\frac{H_{At}}{H_t} = \bar{s}_A \in (0,1), \tag{28}$$

$$x_{it} = \bar{x}_t \equiv \frac{K_t}{A_t} \quad \text{for all } i \in [0, A_t]. \tag{29}$$

As is obvious from the definition, our rule of thumb allocation involves agents in the economy allocating a constant fraction of time to the accumulation of human capital, a constant fraction of output for investment in physical capital, a constant division of human capital into research, and allocating the raw capital symmetrically in the production of the intermediate capital goods.

With this allocation chosen, one can now in principle solve the model for all of the endogenous variables at each point in time. For our purposes, it will be enough to solve for a few key results along the balanced growth path of the economy, which is defined as follows:

DEFINITION 4.2. A *balanced growth path* in this economy is a situation in which all variables grow at constant exponential rates (possibly zero) and in which this constant growth could continue forever.

The following notation will also prove useful in what follows. Let $y \equiv Y/N$ denote final output per capita and let $k \equiv K/N$ represent capital per person. We will use an asterisk superscript to denote variables along a balanced growth path. And finally, g_x will be used to denote the exponential growth rate of some variable x along a balanced growth path.

With this notation, we can now provide a number of useful results for this model.

RESULT 1. With constant allocations of the form given above, this model yields the following results:

(a) Because of the symmetric use of intermediate capital goods, the production function for final output can be written as

$$Y_t = A_t^{\sigma} K_t^{\alpha} H_{Yt}^{1-\alpha}, \quad \sigma \equiv \alpha\left(\frac{1}{\theta} - 1\right). \tag{30}$$

(b) Along a balanced growth path, output per capita y depends on the total stock of ideas, as in

$$y_t^* = \left(\frac{\bar{s}_K}{n + g_k + \delta}\right)^{\alpha/(1-\alpha)} h^*(1 - \bar{s}_A)(1 - \bar{\ell}_h) A_t^{*\sigma/(1-\alpha)}. \tag{31}$$

(c) Along a balanced growth path, the stock of ideas is increasing in the number of researchers, adjusted for their human capital,

$$A_t^* = \left(\frac{\nu}{g_A}\right)^{1/(1-\phi)} H_{At}^{*\lambda/(1-\phi)}. \tag{32}$$

(d) Combining these last two results, output per capita along the balanced growth path is an increasing function of research, which in turn is proportional to the labor force,

$$y_t^* \propto H_{At}^{*\gamma} = (h\bar{s}_A L_t)^{\gamma}, \quad \gamma \equiv \frac{\sigma}{1-\alpha}\frac{\lambda}{1-\phi}. \tag{33}$$

(e) Finally, taking logs and derivatives of these relationships, one gets the growth rates along the balanced growth path,

$$g_y = g_k = \frac{\sigma}{1-\alpha} g_A = \gamma g_{H_A} = g \equiv \gamma n. \tag{34}$$

In general, these results show how the simple model given in the previous section extends when a much richer framework is considered. Result 1(a) shows that this Dixit–Stiglitz technology reduces to a familiar-looking production function when the various capital goods are used symmetrically. Result 1(b) derives the level of output per capita along a balanced growth path, obtaining a solution that is closely related to what one would find in a Solow model. The first term on the right-hand side is simply the capital–output ratio in steady state, the second term adjusts for human capital, the third term adjusts for the fraction of the labor force working to produce goods, and the fourth term adjusts for labor force participation. The final term shows, as in the simple model, that per capita output along a balanced growth path is proportional to the total stock of knowledge (raised to some power).

Result 1(c) provides the analogous expression for the other main production function in the model, the production of ideas. The stock of ideas along a balanced growth path is proportional to the level of the research input (labor adjusted for human capital), again raised to some power. More researchers ultimately mean more ideas in the economy.

Result 1(d) combines these last two expressions to show that per capita output is proportional to the level of research input, which, since human capital per worker is ultimately constant, means that per capita output is proportional to the size of the labor force.[13] The exponent γ essentially measures the total degree of increasing returns to

[13] From now on we will leave the “raised to some power” phrase implicit.

scale in this economy. Notice that it depends on the parameters of both the goods production function and the idea production function, both of which may involve increasing returns.

Finally, Result 1(e) takes logs and derivatives of the relevant "levels" solutions to derive the growth rates of several variables. Output per worker and capital per worker both grow at the same rate. This rate is proportional to the growth rate of the stock of knowledge, which in turn is proportional to the growth rate of the effective level of research. The growth rate of research is ultimately pinned down by the growth rate of population. This last equality parallels the result in the simple model: the fundamental growth rate in the economy is a product of the degree of increasing returns and the rate of population growth. An interesting feature of this result is that the long-run growth rate does not depend on the allocations in this model. Notice that $\bar{s}_A$, for example, does not enter the expression for the long-run growth rate. Changes in the allocation of human capital to research have "level effects", as shown in Result 1(b), but they do not affect the long-run growth rate. This aspect of the model will turn out to be a relatively robust prediction of a class of idea-based growth models.[14]

Pausing to consider the key equations that make up Result 1, the reader might naturally wonder about the restrictive link between the growth rate of human capital and the growth rate of the labor force that has been assumed. For example, in considering Result 1(d), one might accept that per capita output is proportional to research labor adjusted for its human capital, but wonder whether one can get more "action" on the growth side by letting human capital per researcher grow endogenously (in contrast, it is constant in this model).

The answer is that it depends on how one models human capital accumulation. There are many richer specifications of human capital accumulation that deliver results that ultimately resemble those in Result 1. One example is given in footnote 10. Another is given in Chapter 6 of Jones (2002a). In this latter example, an individual's human capital represents the measure of ideas that the individual knows how to work with, which grows over time along a balanced growth path paralleling the growth in knowledge.

An example in which one gets endogenous growth in human capital per worker occurs when one specifies an accumulation equation that is linear in the stock of human capital itself $\dot{h} = \beta e^{\psi \ell_h} h$, reminiscent of Lucas (1988). For reasons discussed in Section 6.2, this approach is unsatisfactory, at least in my view.

4.3. The optimal allocation of resources

The next allocation we will consider is the optimal allocation. That is, we seek to solve for the allocation of resources that maximizes welfare. Because this model is based on a representative agent, this is a straightforward objective, and the optimal allocation is relatively easy to solve for.

[14] This invariance result can be overturned in models in which the population growth rate is an endogenous variable, but the direction of the effects are sometimes odd. See Jones (2003).

DEFINITION 4.3. The *optimal allocation* of resources in this economy consists of time paths $\{c_t, \ell_{ht}, s_{At}, \{x_{it}\}\}_{t=0}^{\infty}$ that maximize utility U_t at each point in time given the economic environment, i.e. given Equations (14)–(25), where $s_{At} \equiv H_{At}/H_t$.

In solving for the optimal allocation of resources, it is convenient to work with the following current-value Hamiltonian

$$\mathcal{H}_t = u(c_t) + \mu_{1t}\big(y_t - c_t - (n+\delta)k_t\big) + \mu_{2t}\nu s_{At}^{\lambda} h_t^{\lambda}(1-\ell_{ht})^{\lambda} N_t^{\lambda} A_t^{\phi}, \tag{35}$$

where

$$y_t = A_t^{\sigma} k_t^{\alpha}\big[(1-s_{At})h_t(1-\ell_{ht})\big]^{1-\alpha}. \tag{36}$$

This last equation incorporates the fact that because of symmetry, the optimal allocation of resources requires the capital goods to be employed in equal quantities.

The current-value Hamiltonian $\mathcal{H}_t$ reflects the utility value of what gets produced at time t: the consumption, the net investment, and the new ideas. As suggested by Weitzman (1976), it is the utility equivalent of net domestic product. The necessary first-order conditions for an optimal allocation can then be written as a set of three control conditions $\partial\mathcal{H}_t/\partial m_t = 0$, where m is a placeholder for c, s_A and ℓ_h and two arbitrage-like equations

$$\bar{\rho} = \frac{\partial\mathcal{H}_t/\partial z_t}{\mu_{it}} + \frac{\dot{\mu}_{it}}{\mu_{it}}, \tag{37}$$

with their corresponding transversality conditions $\lim_{t\to\infty} \mu_{it} e^{-\bar{\rho}t} z_t = 0$. In these expressions, z is a placeholder for k and A, with $i = 1, 2$, respectively, and $\bar{\rho} = \rho - n$ is the effective rate of time preference. The arbitrage interpretation equates the effective rate of time preference to the "dividend" and "capital gain" associated with owning either capital or ideas, where the dividend is the additional flow of utility, $\partial\mathcal{H}_t/\partial z_t$.

RESULT 2. In this economy with the optimal allocation of resources, we have the following results:

(a) All of the results in Result 1 continue to hold, provided the allocations are interpreted as the optimal allocations rather than the rule-of-thumb allocations. For example, output per person along the balanced growth path is proportional to the stock of ideas (raised to some power), which in turn is proportional to the effective amount of research and therefore to the size of the population. As another example, the key growth rates of the economy are determined as in Equation (34), i.e. they are ultimately proportional to the rate of population growth where the factor of proportionality measures the degree of increasing returns in the economy.

(b) The optimal allocation of consumption satisfies the standard Euler equation

$$\frac{\dot{c}_t}{c_t} = \frac{1}{\zeta}\left(\frac{\partial y_t}{\partial k_t} - \delta - \rho\right). \tag{38}$$

(c) The optimal allocation of labor to research equates the value of the marginal product of labor in producing goods to the value of the marginal product of labor in producing new ideas. One way of writing this equation is

$$\frac{s_{At}^{\text{op}}}{1 - s_{At}^{\text{op}}} = \frac{(\mu_{2t}/\mu_{1t})\lambda \dot{A}_t}{(1-\alpha)y_t}, \tag{39}$$

where the "op" superscript denotes the optimal allocation. This equation says that the ratio of labor working to produce ideas to labor working to produce goods is equal to labor's contribution to the value of the new ideas that get produced divided by labor's contribution to the value of output per person that gets produced. Notice that μ_2/μ_1 is essentially the relative price of a new idea in units of output per person.

Along a balanced growth path, we can rewrite this expression as

$$\frac{s_{A}^{\text{op}}}{1 - s_{A}^{\text{op}}} = \frac{\frac{\sigma Y_t/A_t}{r^* - (g_Y - g_A) - \phi g_A}\lambda \dot{A}_t}{(1-\alpha)Y_t}, \tag{40}$$

where $r^* \equiv \rho + \zeta g_c$ functions as the effective interest rate for discounting future output to the present. The relative price of a new idea is given by the presented discounted value of the marginal product of the new idea in the goods production function. This marginal product at one point in time is $\sigma Y/A$, and the equation divides by $r^* - (g_Y - g_A) - \phi g_A$ to adjust for time discounting, growth in this marginal product over time at rate $g_Y - g_A$, and an adjustment for the fact that each new idea helps to produce additional ideas according to the spillover parameter ϕ. Finally, one can cancel the Y's from the numerator and denominator and replace $\dot{A}/A$ by g_A to get a closed-form solution for the allocation of labor to research along a balanced growth path.

(d) The optimal saving rate in this economy along a balanced growth path can be solved for from the Euler equation and the capital accumulation equation. It is given by

$$s_K^{\text{op}} = \frac{\alpha(n+g+\delta)}{\rho + \delta + \zeta g}, \tag{41}$$

where g is the underlying growth rate of the economy, given in Result 1(e). Notice that the optimal investment rate is proportional to the ratio of the marginal product of capital evaluated at the golden rule, $n+g+\delta$, to the marginal product of capital evaluated at the modified golden rule, $\rho + \delta + \zeta g$.

(e) The optimal allocation of time to human capital accumulation is straightforward in this model, and essentially comes down to picking ℓ_h to maximize $e^{\psi \ell_h} \times (1 - \ell_h)$. The solution is to set $\ell_{ht}^{\text{op}} = 1 - 1/\psi$ for all t. As mentioned before, this model introduces human capital in a simple fashion, so the optimal allocation is correspondingly simple.

4.4. A Romer-style equilibrium with imperfect competition

A natural question to ask at this point is whether some kind of market equilibrium can reproduce the optimal allocation of resources. The discussion at the beginning of this chapter made clear the kind of problems that an equilibrium allocation will have to face: the economy is characterized by increasing returns and therefore a standard competitive equilibrium will generally not exist and will certainly not generate the optimal allocation of resources. We are forced to depart from a perfectly competitive economy with no externalities, and therefore one will not be surprised to learn in this section that the equilibrium economy, in the absence of some kind of policy intervention, does not generally reproduce the optimal allocation of resources.

In this section, we study the equilibrium with imperfect competition first described for a model like this by Romer (1990). Romer built on the analysis by Ethier (1982), who extended the consumer variety approach to imperfect competition of Spence (1976) and Dixit and Stiglitz (1977) to the production side of the economy. The economic environment (potentially) involves departures from constant returns in two places, the production function for the consumption–output good and the production function for ideas. We deal with these departures by introducing imperfect competition for the former and externalities for the latter.

Briefly, the economy consists of three sectors. A final goods sector produces the consumption–capital–output good using labor and a collection of capital goods. The capital goods sector produces a variety of different capital goods using ideas and raw capital. Finally, the research sector employs human capital in order to produce new ideas, which in this model are represented by new kinds of capital goods. The final goods sector and the research sector are perfectly competitive and characterized by free entry, while the capital goods sector is the place where imperfect competition is introduced. When a new design for a capital good is discovered, the design is awarded an infinitely-lived patent. The owner of the patent has the exclusive right to produce and sell the particular capital good and therefore acts as a monopolist in competition with the producers of other kinds of capital goods. The monopoly profits that flow to this producer ultimately constitute the compensation to the researchers who discovered the new design in the first place.

As is usually the case, defining the equilibrium allocation of resources in a growth model is more complicated than defining the optimal allocation of resources (if for no other reason than that we have to specify markets and prices). We will begin by stating the key decision problems that have to be solved by the various agents in the economy and then we will put these together in our formal definition of equilibrium.

PROBLEM (HH). Households solve a standard optimization problem, choosing a time path of consumption and an allocation of time. That is, taking the time path of $\{w_t, r_t\}$ as given, they solve

$$\max_{\{c_t,\ell_{ht},\ell_t\}} \int_0^\infty N_t u(c_t) \mathrm{e}^{-\rho t}\, \mathrm{d}t \tag{42}$$

subject to

$$\dot{v}_t = (r_t - n)v_t + w_t h_t \ell_t - c_t, \quad v_0 \text{ given}, \tag{43}$$

$$h_t = \mathrm{e}^{\psi \ell_{ht}}, \tag{44}$$

$$\ell_{ht} + \ell_t = 1, \tag{45}$$

$$N_t = N_0 \mathrm{e}^{nt}, \tag{46}$$

$$\lim_{t\to\infty} v_t \exp\left\{-\int_0^t (r_s - n)\,\mathrm{d}s\right\} \geqslant 0, \tag{47}$$

where v_t is the financial wealth of an individual, w_t is the wage rate per unit of human capital, and r_t is the interest rate.

PROBLEM (FG). A perfectly competitive final goods sector takes the variety of capital goods in existence as given and uses the production technology in Equation (14) to produce output. That is, at each point in time t, taking the wage rate w_t, the measure of capital goods A_t, and the prices of the capital goods p_{it} as given, the representative firm solves

$$\max_{\{x_{it}\}, H_{Yt}} \left(\int_0^{A_t} x_{it}^{\theta}\,\mathrm{d}i\right)^{\alpha/\theta} H_{Yt}^{1-\alpha} - w_t H_{Yt} - \int_0^{A_t} p_{it} x_{it}\,\mathrm{d}i. \tag{48}$$

PROBLEM (CG). Each variety of capital good is produced by a monopolist who owns a patent for the good, purchased at a one-time price P_{At}. As discussed in describing the economic environment, one unit of the capital good can be produced with one unit of raw capital. The monopolist sees a downward-sloping demand curve for her product from the final goods sector and chooses a price to maximize profits. That is, at each point in time and for each capital good i, a monopolist solves

$$\max_{p_{it}} \pi_{it} \equiv (p_{it} - r_t - \delta)x(p_{it}), \tag{49}$$

where $x(p_{it})$ is the demand from the final goods sector for intermediate good i if the price is p_{it}. This demand curve comes from a first-order condition in Problem (FG). The monopoly profits are the revenue from sales of the capital goods less the cost of the capital need to produce the capital goods (including depreciation). The monopolist is small relative to the economy and therefore takes aggregate variables and the interest rate r_t as given.[15]

[15] To be more specific, the demand curve $x(p_i)$ is given by

$$x(p_{it}) = \left(\alpha \frac{Y}{\int_0^{A_t} x_{it}^{\theta}\,\mathrm{d}i} \frac{1}{p_{it}}\right)^{1/(1-\theta)}.$$

We assume the monopolist is small relative to the aggregate so that it takes the price elasticity to be $-1/(1-\theta)$.

PROBLEM (R&D). The research sector produces ideas according to the production function in Equation (16). However, each individual researcher is small and takes the productivity of the idea production function as given. In particular, each researcher assumes that the idea production function is

$$\dot{A}_t = \bar{\nu}_t H_{At}. \tag{50}$$

That is, the duplication effects associated with λ and the knowledge spillovers associated with ϕ in Equation (16) are assumed to be external to the individual researcher. In this perfectly competitive research sector, the representative research firm solves

$$\max_{H_{At}} P_{At}\bar{\nu}_t H_{At} - w_t H_{At}, \tag{51}$$

taking the price of ideas P_{At}, research productivity $\bar{\nu}_t$, and the wage rate w_t as given.

Now that these decision problems have been described, we are ready to define an equilibrium with imperfect competition for this economy.

DEFINITION 4.4. An *equilibrium with imperfect competition* in this economy consists of time paths for the allocations $\{c_t, \ell_{ht}, \ell_t, \{x_{it}\}, Y_t, K_t, v_t, \{\pi_{it}\}, H_{Yt}, H_{At}, H_t, h_t, L_t, N_t, A_t, \bar{\nu}_t\}_{t=0}^{\infty}$ and prices $\{w_t, r_t, \{p_{it}\}, P_{At}\}_{t=0}^{\infty}$ such that for all t:

1. c_t, v_t, h_t, ℓ_{ht} and ℓ_t solve Problem (HH).
2. $\{x_{it}\}$ and H_{Yt} solve Problem (FG).
3. p_{it} and π_{it} solve Problem (CG) for all $i \in [0, A_t]$.
4. H_{At} solves Problem (R&D).
5. (r_t) The capital market clears: $V_t \equiv v_t N_t = K_t + P_{At}A_t$.
6. (w_t) The labor market clears: $H_{Yt} + H_{At} = H_t$.
7. $(\bar{\nu}_t)$ The idea production function is satisfied: $\bar{\nu}_t = \nu H_{At}^{\lambda-1} A_t^{\phi}$.
8. (K_t) The capital resource constraint is satisfied: $\int_0^{A_t} x_{it}\,\mathrm{d}i = K_t$.
9. (P_{At}) Assets have equal returns: $r_t = \frac{\pi_{it}}{P_{At}} + \frac{\dot{P}_{At}}{P_{At}}$.
10. Y_t is given by the production function in (14).
11. A_t is given by the production function in (16).
12. $H_t = h_t L_t$.
13. $L_t = \ell_t N_t$ and $N_t = N_0 \mathrm{e}^{nt}$.

Notice that, roughly speaking, there are twenty equilibrium objects that are part of the definition of equilibrium and there are twenty equations described in the conditions for equilibrium that determine these objects at each point in time.[16] Not surprisingly, one cannot solve in general for the equilibrium outside of the balanced growth path, but along a balanced growth path the solution is relatively straightforward, and we have the following results.

[16] The condition omitted from this definition of equilibrium is the law of motion for the capital stock, given in the economic environment in Equation (15). That this equation holds in equilibrium is an implication of Walras' law. It can be derived in equilibrium by differentiating the capital market clearing condition that $V = K + P_A A$ with respect to time and making the natural substitutions.

RESULT 3. In the equilibrium with imperfect competition:

(a) All of the results in Result 1 continue to hold along the balanced growth path, provided the allocations are interpreted as the equilibrium allocations rather than the rule-of-thumb allocations. For example, output per person along the balanced growth path is proportional to the stock of ideas (raised to some power), which in turn is proportional to the effective amount of research and therefore to the size of the population. As another example, the key growth rates of the economy are determined as in Equation (34), i.e. they are ultimately proportional to the rate of population growth where the factor of proportionality measures the degree of increasing returns in the economy.

(b) The Euler equation for consumption and the allocation of time to human capital accumulation are undistorted in this equilibrium. That is, the equations that apply are identical to the equations describing the optimal allocation of resources:

$$\frac{\dot{c}_t}{c_t} = \frac{1}{\zeta}\left(r_t^{\mathrm{eq}} - \rho\right), \tag{52}$$

$$\ell_{ht}^{\mathrm{eq}} = 1 - \frac{1}{\psi}. \tag{53}$$

(c) The solution to Problem (CG) involves a monopoly markup over marginal cost that depends on the CES parameter in the usual way,

$$p_{it}^{\mathrm{eq}} = p_t^{\mathrm{eq}} \equiv \frac{1}{\theta}\left(r_t^{\mathrm{eq}} + \delta\right). \tag{54}$$

Because of this monopoly markup, however, capital is paid less than its marginal product, and the equilibrium interest rate is given by

$$r_t^{\mathrm{eq}} = \alpha\theta\frac{Y_t}{K_t} - \delta. \tag{55}$$

Because the equilibrium economy grows at the same rate as the economy with optimal allocations, the steady-state interest rate determined from the Euler equation is the same in the two economies. Therefore, the fact that capital is paid less than its marginal product translates into a suboptimally low capital–output ratio in the equilibrium economy. Similarly, the equilibrium investment rate along a balanced growth path is given by

$$s_K^{\mathrm{eq}} = \frac{\alpha\theta(n+g+\delta)}{\rho+\delta+\zeta g} = \theta s_K^{\mathrm{op}}. \tag{56}$$

(d) The equilibrium allocation of human capital to research equates the wage of human capital in producing goods to its wage in producing ideas. This result can be written in an equation analogous to (39) as

$$\frac{s_{At}^{\mathrm{eq}}}{1 - s_{At}^{\mathrm{eq}}} = \frac{P_{At}\dot{A}_t}{(1-\alpha)Y_t}. \tag{57}$$

The ratio of the share of human capital working to produce ideas to that working to produce goods is equal to the value of the output of new ideas divided by labor's share of the value of final goods.

Along the balanced growth path, we can rewrite this expression as

$$\frac{s_A^{\mathrm{eq}}}{1-s_A^{\mathrm{eq}}} = \frac{\frac{\sigma\theta Y_t/A_t}{r^{\mathrm{eq}}-(g_Y-g_A)}\dot{A}_t}{(1-\alpha)Y_t}, \tag{58}$$

which is directly comparable to the optimal allocation in Equation (40). In comparing these two equations, we see three differences. The first two differences reflect the externalities in the idea production function. The true marginal product of human capital in research is lower by a factor of $\lambda < 1$ than the equilibrium economy recognizes because of the congestion/duplication externality, which tends to lead the equilibrium to overinvest in research. On the other hand, the equilibrium allocation ignores the fact that the discovery of new ideas may raise the future productivity of research if $\phi > 0$. This changes the effective rate at which the flow of future ideas is discounted, potentially causing the equilibrium to underinvest in research. Finally, the third difference reflects the appropriability of returns. A new idea raises the current level of output in the final goods sector according to the marginal product $\sigma Y/A$. However, the research sector appropriates only the fraction $\theta < 1$ of this marginal product. The reason is familiar from the standard monopoly diagram in undergraduate classes: the profits appropriated by a monopolist are strictly lower than the consumer surplus created by that monopolist. This appropriability effect works to cause the equilibrium allocation of human capital to research to be too low. Overall, these three distortions do not all work in the same direction, so that theory cannot tell us whether the equilibrium allocation to research is too high or too low.

4.5. Discussion

Let us step back for a moment to take stock of what we learn from the developments in this section. The most important finding is Result 1, together with the fact that it carries over into the other allocations as Result 2(a) and Result 3(a). This result is simply a confirmation of the basic results from the simple model in Section 3. Because of the nonrivalrous nature of ideas, output per person depends on the total stock of ideas in the economy instead of the per capita stock of ideas. This is a direct implication of the fact that nonrivalry leads to increasing returns to scale. In turn, it implies that output per capita, in the long run, is an increasing function of the total amount of research, which in turn is an increasing function of the scale of the economy, measured by the size of its total population. Log-differencing this statement, we see that the growth rate of output per worker ultimately depends on the growth rate of the number of researchers and therefore on the growth rate of population. This has been analyzed and discussed extensively in a number of recent papers; these will be reviewed in detail in Section 5.

The second main finding from models like this is that the equilibrium allocation of resources is not generally optimal, at least not in the absence of some kind of policy intervention. Here, the allocation of resources to the production of new ideas can be either too high or too low, as discussed above.[17] In addition, investment rates are too low in equilibrium, reflecting the fact that capital is paid less than its marginal product so that some resources are available to compensate inventive effort.

In this equilibrium, the suboptimal allocation of resources is easily remedied. A subsidy to capital accumulation and a subsidy or tax on research can be financed with lump sum taxes in order to generate the optimal allocation of resources. A useful exercise is to solve for the equilibrium in the presence of such taxes in order to determine the optimal tax rates along a balanced growth path.

Given the simplicity of this economic environment, there exist alternative institutions that are equally effective in getting optimal allocations. For example, consider a perfectly competitive economy in which all research is publicly-funded. The government raises revenue with lump-sum taxes and uses these taxes to hire researchers that produce new ideas. These new ideas are then released into the public domain where anyone can use them to produce capital goods in perfect competition.[18]

In practice of course, one suspects that obtaining the optimal allocation of resources is more difficult than either the world of imperfect competition with taxes and subsidies or the perfectly-competitive world with public funding of research suggest. There are many different directions for research, many different kinds of labor (different skill levels and talents), and individual effort choices that are unobserved by the government. Indeed, the available evidence suggests that the allocation of resources to research falls short of the optimal level. Jones and Williams (1998) take advantage of a large body of empirical work in the productivity literature to conclude that the social rate of return to research substantially exceeds the private rate of return, suggesting that research effort falls short of the optimum.

The implication of this is that there is no reason to think that we have found the best institutions for generating the optimal allocation of resources to research. Institutions like the patent system or the Small Business Innovative Research (SBIR) grants program are themselves ideas. These institutions have evolved over time to promote an efficient

[17] This conclusion also holds true in the Schumpeterian growth models of Aghion and Howitt (1992) and Grossman and Helpman (1991) discussed in Chapter 2 of this Handbook, but for a different reason. In these quality-ladder models, a firm discovers a better version of an existing product, displacing the incumbent producer. Some of the rents earned by the innovator are the result of past discoveries, and some of the rents earned by future innovators will be due to the discovery of the current innovator. This business stealing creates another distortion in the allocation of resources to research. Because an innovator essentially steals first and gets expropriated later, the effect of this business stealing distortion is to promote excessive research. Because the model also features appropriability problems and knowledge spillovers, the equilibrium amount of research can be either too high or too low in these models.

[18] A useful exercise here is to define the competitive equilibrium with public funding of research and to solve for optimal taxes and public expenditure.

allocation of resources, but it is almost surely the case that better institutions – better ideas – are out there to be discovered.

Interestingly, this result can be illustrated within the model itself. Notice how much easier it is to define the optimal allocation than it is to define the equilibrium allocation. The equilibrium with imperfect competition requires the modeler to be "clever" and to come up with the right institutions (e.g., a patent system, monopolistic competition, and the appropriate taxes and subsidies) to make everything work out. In reality, society must invent and implement these institutions.

Three recent papers deserve mention in this context. Romer (2000) argues that subsidizing the key input into the production of ideas – human capital in the form of college graduates with degrees in engineering and the natural sciences – is preferable to government subsidies downstream like the SBIR program. Kremer (1998) notes the large ex-post monopoly distortions associated with patents in the pharmaceutical industry and elsewhere and proposes a new mechanism for encouraging innovation. In particular, he suggests that the government (or other altruistic organizations such as charitable foundations) should consider purchasing the patents for particular innovations and releasing them into the public domain to eliminate the monopoly distortion. Boldrin and Levine (2002), in a controversial paper, are even more critical of existing patent and copyright systems and propose restricting them severely or even eliminating them altogether.[19] They argue that first-mover advantages, secrecy, and imitation delays provide ample protection for innovators and that an economy without patent and copyright systems would have a better allocation of resources than the current regime in which copyright protection is essentially indefinite and patents are used as a weapon to discourage innovation. Each of these papers makes a useful contribution by attempting to create new institutions that might improve the allocation of resources.

5. Scale effects

Idea-based growth models are linked tightly to increasing returns to scale, as was noted earlier in the Idea Diagram. The mechanism at the heart of this link is nonrivalry: the fact that knowledge can be used by an arbitrarily large number of people simultaneously without degradation means that there is something special about the first instantiation of an idea. There is a cost to creating an idea in the first place that does not have to be re-incurred as the idea gets used by more and more people. This fixed cost implies that production is, at least in the absence of some other fixed factor like land, characterized by increasing returns to scale.

Notice that nothing in this argument relies on a low marginal cost of production or on the absence of learning and human capital. Consider the design of a new drug for treating high blood pressure. Discovering the precise chemical formulation for the drug

[19] See also the important elaborations and clarifications in Quah (2002).

may require hundreds of millions of dollars of research effort. This idea is then simply a chemical formula. Producing copies of the drug – pills – may be expensive, for example if the drug involves the use of a rare chemical compound. It may also be such that only the best-trained biochemists have the knowledge to understand the chemical formula and manufacture the drug. Nevertheless, an accurate characterization of the production technology for producing the drug is as a fixed research cost followed by a constant marginal cost. Once the chemical formula is discovered, to double the production of pills we simply double the number of highly-trained biochemists, build a new (identical) factory, and purchase twice as much of the rare chemical compound used as an input.

Because the link between idea-based growth theory and increasing returns is so strong, the role of "scale effects" in growth models has been the focus of a series of theoretical and empirical papers. In discussing these papers, it is helpful to consider two forms of scale effects. In models that exhibit "strong" scale effects, the growth rate of the economy is an increasing function of scale (which typically means overall population or the population of educated workers). Examples of such models include the first-generation models of Romer (1990), Aghion and Howitt (1992) and Grossman and Helpman (1991). On the other hand, in models that exhibit "weak" scale effects, the level of per capita income in the long run is an increasing function of the size of the economy. This is true in the "semi-endogenous" growth models of Jones (1995a), Kortum (1997) and Segerstrom (1998) that were written at least partially in response to the strong scale effects in the first generation models. The models examined formally in the previous sections of this chapter fit into this category as well.

To use an analogy from the computer software industry, are scale effects a bug or a feature? I believe the correct answer is slightly complicated. I will argue that overall they are a feature, i.e. a useful prediction of the model that helps us to understand the world. However, in some papers, most notably in the first generation of idea-based growth models, these scale effects appeared in an especially potent way, producing predictions in these models that are easily falsified. This strong form of scale effects – in which the long-run growth rate of the economy depends on its scale – is a bug. Subsequent research has remedied this problem, maintaining everything that is important about idea-based growth models but eliminating the strong form of the scale effects prediction. This still leaves us, as discussed above, with a weak form of scale effects: the size of the economy affects, in some sense, the level of per capita income. This, of course, is nothing more than a statement that the economy is characterized by increasing returns to scale. The weak form of scale effects has its critics as well, but I will argue two things. First, these criticisms are generally misplaced. And second, it's fortunate that this is the case: the weak form of scale effects is so inextricably tied to idea-based growth models that rejecting one is largely equivalent to rejecting the other.

The remainder of this section consists of two basic parts. Section 5.1 returns to the simple growth model presented in Section 3 to formalize the strong and weak versions of scale effects. The remaining sections then discuss a range of applications in the literature related to scale effects.

5.1. Strong and weak scale effects

The simple model in Section 3 revealed that the growth rate of per capita income is proportional to the growth rate of the stock of ideas. Consider that same model, but replace the idea production function in Equation (7) with

$$\dot{A}_t = \nu L_{At}^{\lambda} A_t^{\phi}. \tag{59}$$

We could go further and incorporate human capital, as we did in the richer model of Section 4, but this will not change the basic result, so we will leave out this complication.

Now consider two cases. In the first, we impose the condition that $\phi < 1$. In the second, we will instead assume that $\phi = 1$. In the case of $\phi < 1$, the analysis goes through exactly as in the models developed earlier, and the growth rate of the stock of ideas along a balanced growth path is given by

$$g_A = \frac{\lambda n}{1-\phi}, \tag{60}$$

which pins down all the key growth rates in the model. Notice that, as before, the growth rate is proportional to the rate of population growth. It is straightforward to show, as we did earlier, that the level of per capita income in such an economy is an increasing function of the size of the population. That is, this model exhibits weak scale effects. Finally, notice that this equation cannot apply if $\phi = 1$; in that case, the denominator would explode.

To see more clearly the source of the problem, rewrite the idea production function when we assume $\phi = 1$ as

$$\frac{\dot{A}_t}{A_t} = \nu L_{At}^{\lambda}. \tag{61}$$

In this case, the growth rate of knowledge is proportional to the number of researchers raised to some power λ. If the number of researchers is itself growing over time, the simple model will not exhibit a balanced growth path. Rather, the growth rate itself will be growing! With $\phi = 1$, the simple model exhibits strong scale effects.

The first generation idea-based growth models of Romer (1990), Aghion and Howitt (1992) and Grossman and Helpman (1991) all include idea production functions that essentially make the assumption of $\phi = 1$, and all exhibit the strong form of scale effects.[20] The problem with the strong form of scale effects is easy to document and understand. Because the growth rate of the economy is an increasing function of research effort, these models require research effort to be constant over time to match the relative

[20] This is easily seen in the Romer expanding variety model, as that model is the building block for the models developed in this chapter. It is slightly trickier to see this in the quality ladder models of Aghion and Howitt (1992) and Grossman and Helpman (1991). In those models, each researcher produces a constant number of ideas, but ideas get bigger over time. In particular, each new idea generates a *proportional* improvement in productivity.

stability of growth rates in the United States and some other advanced economies. However, research effort is itself growing over time (for example, if for no other reason than simply because the population is growing). These facts are now documented in more detail.

A useful stylized fact that any growth model must come to terms with is the relative stability of growth rates in the United States over more than a century. This stability can be easily seen by plotting per capita GDP for the United States on a logarithmic scale, as shown in Figure 1. A straight line with a growth rate of 1.8 percent per year provides a very accurate description of average growth rates in the United States dating back to 1870. There are departures from this line, of course, most clearly corresponding to the Great Depression and the recovery following World War II. But what is truly remarkable about this figure is how well a straight line describes the trend.

Jones (1995b) made this point in the following way. Suppose one drew a trend line using data from 1870 to 1929 and then extrapolated that line forward to predict per capita GDP today. It turns out that such a prediction matches up very well with the current level of per capita GDP, confirming the hypothesis that growth rates have been relatively stable on average.[21]

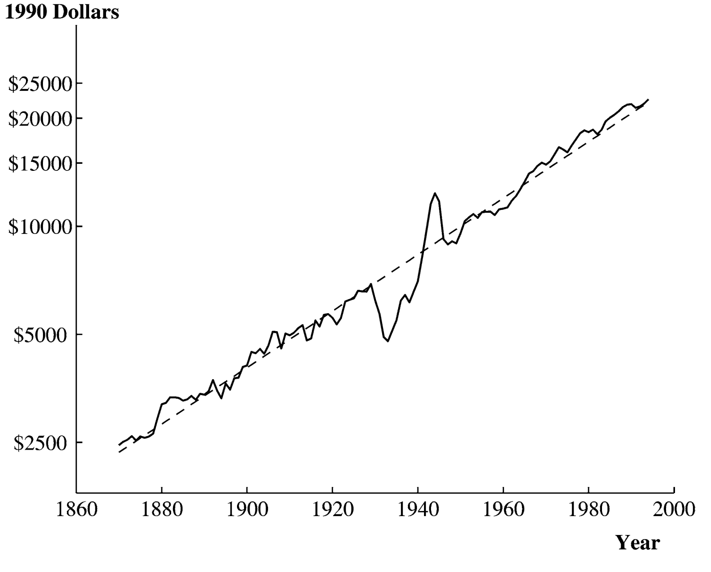

Figure 1. U.S. GDP per capita log scale. *Source*: Maddison (1995).

[21] Of course this is only an approximation. The growth rate from 1950 to 1994 averaged 1.95 percent, while the growth rate from 1870 to 1929 averaged 1.75 percent [see, e.g., Ben-David and Papell (1995) on this increase]. On the other hand, the 2.20 percent growth rate in the 1950s and 1960s is slightly higher than the 1.74 percent growth rate after 1970, reflecting the productivity slowdown. Similar results can be obtained with GDP per worker and GDP per hour worked, see Williams (1995).

This stylized fact represents an important benchmark that any growth model must match. Whatever the engine driving long run growth, it must (a) be able to produce relatively stable growth rates for a century or more, and (b) must not predict that growth rates in the United States over this period of time should depart from such a pattern. To see this force of this argument, consider first a theory like Lucas (1988) that predicts that investment in human capital is the key to growth. In this model, the growth rate of the economy is proportional to the investment rate in human capital. But if investment rates in human capital have risen significantly in the 20th century in the United States, as data on educational attainment suggests, this is a problem for the theory. It could be rescued if investment rates in human capital in the form of on-the-job training have fallen to offset the rise in formal education, but there is little evidence suggesting that this is the case.

This stylized fact is even more problematic for the first-generation idea-based growth models of Romer (1990), Aghion and Howitt (1992) and Grossman and Helpman (1991) (R/AH/GH). These models predict that growth is an increasing function of research effort, but research effort has apparently grown tremendously over time. As one example of this fact, consider Figure 2. This figure plots an index of the number of scientists and engineers engaged in research in the G-5 countries. Between 1950 and 1993, this index of research effort rose by more than a factor of eight. In part this is because of the general growth in employment in these countries, but as the figure shows, it also reflects a large increase in the fraction of employment devoted to research. A similar fact can be

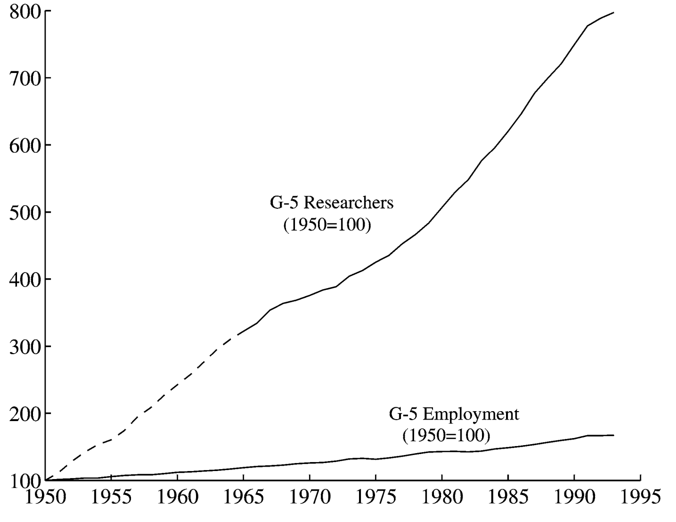

Figure 2. Researchers and employment in the G-5 countries (index). *Note*. From calculations in Jones (2002b). Data on researchers before 1950 in countries other than the United States is backcasted using the 1965 research share of employment. The G-5 countries are France, Germany, Japan, the United Kingdom and the United States.

documented using just the data for the United States, or by looking at spending on R&D rather than employment.[22] The bottom line is that resources devoted to research have exhibited a tremendous amount of growth in the post-war period, while growth rates in the United States have been relatively stable. The implication is that models that exhibit strong scale effects are inconsistent with the basic trends in aggregate data. Evidence like this is one of the main arguments in favor of models that exhibit weak scale effects instead.[23]

5.2. *Growth effects and policy invariance*

At some level, the rejection of models with strong scale effects in favor of models with weak scale effects should not be especially interesting. The only difference between the two models, as discussed above, is essentially the strength of the knowledge spillover parameter. In expanding variety models, is $\phi = 1$ or is $\phi < 1$? Nothing in the evidence necessarily rules out $\phi = 0.95$, and continuity arguments suggest that the economics of $\phi = 0.95$ and $\phi = 1$ cannot be that different.

The main difference in the economic results that one obtains in the two models pertains to the ability of changes in policy to alter the long-run growth rate of the economy. In the models that exhibit strong scale effects, the long-run growth rate is an increasing function of the number of researchers. Hence, a policy that increases the number of researchers, such as an R&D expenditure subsidy, will increase the long-run growth rate. In contrast, if $\phi < 1$, then the long-run growth rate depends on elasticities of production functions and on the rate of population growth. To the extent that these parameters are unaffected by policy – as one might naturally take to be the case, at least to a first approximation – policy changes such as a subsidy to R&D or a tax on capital will have no affect on the long-run growth rate. They will of course affect the long-run level of income and will affect the growth rate along a transition path, but the long-run growth rate is invariant to standard policy changes.

This statement can be qualified in a couple of ways. First, the population growth rate is really an endogenous variable determined by fertility choices of individuals. Policy changes can affect this choice and hence can affect long-run growth even in a model with weak scale effects, as shown in Jones (2003). However, the effects can often be counter to the usual direction. For example, a subsidy to R&D can lead people to perform more research and have fewer kids, reducing fertility. Hence a subsidy to research

[22] There are several ways to look at the R&D spending share of GDP. For total R&D expenditures as a share of GDP in the United States, most of the increase in the R&D share occurs before 1960. However, if one substracts out R&D expenditures on defense and space (which might be a reasonable thing to do since government output is valued at cost), or if one focuses on non-federally-financed research, the trend in the U.S. share emerges clearly; see Chapter 4 of the NSF's *Science and Engineering Indicators, 2004*. Alternatively, there are substantial trends in the R&D shares for most of the other G-7 countries; in addition to the 2004 edition, see also the 1993 edition of the NSF's *Science and Engineering Indicators* to get data on the research shares back to 1970.

[23] The other main argument is the "linearity critique" discussed further in Section 6.2.

can reduce the long-run growth rate. This can be true even if it is optimal to subsidize research – this kind of model makes clear that long-run growth and welfare are two very different concepts. The second qualification is that one can imagine subsidies that affect the direction of research and that can possibly affect long-run growth. For example, Cozzi (1997) constructs a model in which research can proceed in different directions that may involve different knowledge spillover elasticities. By shifting research to the directions with high spillovers, it is possible to change the long-run growth rate.

Despite these qualifications, it remains true that in the semi-endogenous growth models written to address the problem of strong scale effects, straightforward policies do not affect the long-run growth rate. This has led a number of researchers to seek alternative means of eliminating the strong scale effects prediction while maintaining the potency of policy to alter the long-run growth rate. Key papers in this line of research include Young (1998), Peretto (1998), Dinopoulos and Thompson (1998) and Howitt (1999) (Y/P/DT/H).

These papers all work in a similar way.[24] In particular, each adds a second dimension to the model, so that research can improve productivity for a particular product or can add to the variety of products. To do this in the simplest way, suppose that aggregate consumption (or output) is a CES composite of a variety of different products

$$C_t = \left(\int_0^{B_t} Y_{it}^{1/\theta} \, \mathrm{d}i \right)^{\theta}, \quad \theta > 1, \tag{62}$$

where B_t represents the variety of goods that are available at date t and Y_{it} is the output of variety i. Assume that each variety Y_i is produced using the Romer-style technology with $\phi = 1$ in the simple model given earlier in Section 3.

The key to the model is the way in which the number of different varieties B changes over time. To keep the model as simple as possible, assume

$$B_t = L_t^{\beta}. \tag{63}$$

That is, the number of varieties is proportional to the population raised to some power β. Notice that this relationship could be given microfoundations with an idea production function analogous to that in Equation (16).[25]

Finally, let us assume each intermediate variety is used in the same quantity so that $Y_{it} = Y_t$, implying $C_t = B_t^{\theta} Y_t$. Per capita consumption is then $c_t = B_t^{\theta} y_t$, and per capita consumption growth along a balanced growth path is

$$g_c = \theta g_B + \sigma g_A = \theta \beta n + \sigma g_A. \tag{64}$$

Assuming an idea production function with $\phi = 1$, like that in R/AH/GH, the growth rate of the stock of ideas is proportional to research effort per variety, $L_{At}/B_t = sL_t/B_t$,

$$g_A = \frac{\nu s L_t}{B_t} = \nu s L_t^{1-\beta}. \tag{65}$$

[24] This section draws heavily on Jones (1999).

[25] For example, if $\dot{B} = LB^{\gamma}$, then Equation (63) holds along a balanced growth path with $\beta = 1/(1 - \gamma)$.

Substituting this equation back into (64) gives the growth rate of per capita output as a function of exogenous variables and parameters

$$g_c = \theta\beta n + \sigma\nu s L_t^{1-\beta}. \tag{66}$$

With $\beta = 1$ so that $B_t = L_t$, the strong scale effect is eliminated from the model, while the effect of policy on long-run growth is preserved. That is, a permanent increase in the fraction of the labor force working in research, s, will permanently raise the growth rate. This is the key result sought by the Y/P/DT/H models.

However, there are two important things to note about this result. First, it is very fragile. In particular, to the extent that $\beta \neq 1$, problems reemerge. If $\beta < 1$, then the model once again exhibits strong scale effects. Alternatively, if $\beta > 1$, then changes in s no longer permanently affect the long-run growth rate. Thus, the Y/P/DT/H result depends crucially on a knife-edge case for this parameter value, in addition to the Romer-like knife-edge assumption of $\phi = 1$. Second, as the first term in Equation (66) indicates, the model still exhibits the weak form of scale effects. This result is not surprising given that these are idea-based growth models, but it is useful to recognize since many of the papers in this literature have titles that include the phrase "growth without scale effects". What these titles really mean is that the papers attempt to eliminate strong scale effects; all of them still possess weak scale effects. These points are discussed in more detail in Jones (1999) and Li (2000, 2002).

5.3. *Cross-country evidence on scale effects*

One source of evidence on the empirical relevance of scale effects comes from looking across countries or regions at a point in time. Consider first the ideal cross-sectional evidence. One would observe two regions, one larger than the other, that are otherwise identical. The two regions would not interact in any way and the only source of new ideas in the two regions would be the regions' own populations. In such an ideal experiment, one could search for scale effects by looking at the stock of ideas and at per capita income in each region over time. In the long-run, one would expect that the larger region would end up being richer.

In practice, of course, this ideal experiment is never observed. Instead, we have data on different countries and regions in the world, but these regions almost certainly share ideas and they almost certainly are not equal in other dimensions. It falls to clever econometricians to use this data to approximate the ideal experiment. No individual piece of evidence is especially compelling, but the collection taken together does indeed suggest that the cross-sectional evidence on scale effects supports the basic model.

Certainly the most creative approximation to date is found in Kremer (1993) and later appears in the Pulitzer Prize-winning book *Guns, Germs, and Steel* by Diamond (1997). The most recent ice age ended about 10,000 B.C. Before that time, ocean levels were lower, allowing humans to migrate around the world – for example across the Bering Strait and into the Americas. In this sense, ideas could diffuse across regions. However, with the end of the ice age, sea levels rose, and various regions of the world were effectively isolated from each other, at least until the advent of large sailing ships sometime around the year 1000 or 1500. In particular, for approximately 12,000 years, five regions were mutually isolated from one another: the Eurasian/African continents, the Americas, Australia, Tasmania (an island off the coast of Australia), and the Flinders Island (a very small island off the coast of Tasmania). These regions are also nicely ranked in terms of population sizes, from the relatively highly-populated Eurasian/African continent down to the small Flinders Island, with a population that likely numbered fewer than 500.

It is plausible that 12,000 years ago these regions all had similar technologies: all were relatively primitive hunter-gatherer cultures. Now fast-forward to the year 1500 when a wave of European exploration reintegrates the world. First, the populous Old World has the highest level of technological sophistication; they are the ones doing the exploring. The Americas follow next, with cities, agriculture, and the Aztec and Mayan civilizations. Australia is in the intermediate position, having developed the boomerang, the atlatl, fire-making, and sophisticated stone tools, but still consisting of a hunter-gatherer culture. Tasmania is relatively unchanged, and the population of Flinders Island had died out completely. The technological rank of these regions more than 10,000 years later matches up exactly with their initial population ranks at the end of the last ice age.

Turning to more standard evidence from the second-half of the 20th century, one is first struck by the apparent lack of support for the hypothesis of weak scale effects. The most populous countries of the world, China and India, are among the poorest, while some of the smallest countries like Hong Kong and Luxembourg are among the richest. And the countries with the most rapid rates of population growth – many in Africa – are among the countries with the slowest rates of per capita income growth. However, a moment's thought suggests that one must be careful in interpreting this evidence. It is clearly not the case that Hong Kong and Luxembourg are isolated countries that grow solely based on the ideas created by their own populations. These countries benefit tremendously from ideas created around the world. And in the case of the poor countries of the world, "other things" are clearly not equal. These countries have very different levels of human capital and different policies, institutions, and property rights that contribute to their poverty. Hence, we must turn to econometric evidence that seeks to neutralize these differences.

The clearest cross-country evidence in favor of weak scale effects comes from papers that explicitly control for differences in international trade. Intuitively, openness to international trade is likely related to openness to idea flows, and the flow of ideas from other countries is one of the key factors that needs to be neutralized. Backus, Kehoe and Kehoe (1992), Frankel and Romer (1999) and Alcala and Ciccone (2002) are the main examples of this line of work, and all find an important role for scale. Alcala and Ciccone (2002) provide what is probably the best specification, controlling for both trade and institutional quality (and instrumenting for these endogenous variables), but the results in Frankel and Romer (1999) are similar. Alcala and Ciccone find a long-run elasticity of GDP per worker with respect to the size of the workforce that is equal to 0.20.[26] That is, holding other things equal, a 10 percent increase in the size of the workforce in the long run is associated with a 2 percent higher GDP per worker.[27]

Other cross-country studies, of course, have not been able to precisely estimate this elasticity. Hall and Jones (1999), for example, found a point estimate of about 0.05, but with a standard error of 0.06. Sala-i-Martin (1997) does not find the size of the population to be a robust variable in his four million permutations of cross-country growth regressions. Finally, it should be recognized that this cross-country estimate of the scale elasticity is not necessarily an estimate of the structural parameter γ in the idea models presented earlier in Section 4. One needs a theory of technology adoption and idea flows in order to make sense of the estimates. For example, in a world where ideas flow to all places instantaneously, there would be no reason to find a scale effect in the cross-section evidence.

A final piece of evidence that is often misinterpreted as providing evidence against the weak scale effects prediction is the negative coefficient on population growth in a cross-country growth regression, such as in Mankiw, Romer and Weil (1992). Recall that the standard interpretation of these regressions is that they are estimating transition dynamics. The negative coefficient on population growth is interpreted as capturing the dilution of the investment rate associated with the Solow model. Consider two countries that are identical but for different population growth rates. The country with the faster population growth rate must equip a larger number of new workers with the existing capital–labor ratio, effectively diluting the investment rate. The result is that such an economy has a lower capital–output ratio in steady state, reducing output per worker along the balanced growth path. But this same force is also at work in any growth model, including idea-based models, as was apparent above in Result 1(b). The implication is that this cross-country evidence is not inconsistent with models in which weak scale effects play a role.

5.4. Growth over the very long run

Additional evidence on the potential relevance of scale effects to economic growth comes from what at first might seem an unlikely place: the history of growth from thousands of years ago to the present.

One of the important applications of models of economic growth in recent years has been to understand economic growth over this very long time period. Many of our workhorse models of growth were constructed with an eye toward 20th-century growth. Asking how well they explain growth over a much longer period of time therefore provides a nice test of our models.

[26] The standard error of this particular point estimate is about 0.10. Across different specifications, the elasticity ranges from a low of about 0.10 to a high of about 0.40.

[27] Of course in the model with trade, other things would not be equal: a change in population would almost surely affect the trade-GDP ratios that measure openness in the regression.

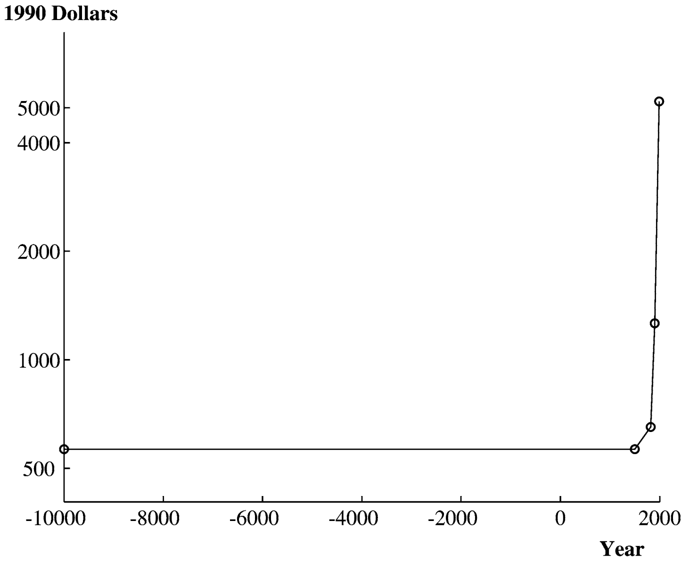

Figure 3. World per capita GDP (log scale). *Note*. Data from Maddison (1995) for years after 1500. Before 1500, we assume a zero growth rate, as suggested by Maddison and others.

The key fact that must be explained over this period is quite stunning and is displayed in Figure 3. For thousands and thousands of years prior to the Industrial Revolution, standards of living were relatively low. In particular, the evidence suggests that there was no sustained growth in per capita incomes before the Industrial Revolution.[28] Then, quite suddenly from the standpoint of the sweep of world history, growth rates accelerated and standards of living began rising with increasing rapidity. At the world level, per capita income today is probably about 10 times higher than it was in the year 1800 or 1500 or even 10,000 years ago. A profound question in economic history – and one that growth economists have begun delving into – is this: How do we understand this entire time path? Why were standards of living relatively low and stagnant for so long, why have they risen so dramatically in the last 150 years, and what changed?[29]

The recent growth literature on this question is quite large, and a thorough review is beyond the scope of the present chapter (additional discussion can be found in Chapter 4 of this Handbook, by Oded Galor). Representative papers include Lee (1988), Kremer

[28] See Lucas (1998), Galor and Weil (2000), Jones (2001) and Clark (2001) for a discussion of this evidence.

[29] A cottage industry (!) in recent years has sprung up in which macroeconomists bring their modeling tools to bear on major questions in economic history. In addition to growth over the very long run, macroeconomists have studied the Great Depression [Ohanian and Cole (2001)], the Second Industrial Revolution [Atkeson and Kehoe (2002)], and the rise in female labor force participation over the course of the 20th century [Greenwood, Seshadri and Yorukoglu (2001)] among other topics.

(1993), Goodfriend and McDermott (1995), Lucas (1998), Galor and Weil (2000), Clark (2001), Jones (2001), Stokey (2001), Hansen and Prescott (2002) and Tamura (2002).

In several of these papers, scale effects play a crucial role. Scale is at the heart of the models of Lee (1988), Kremer (1993) and Jones (2001), and it also plays an important role in getting growth started in the model based on human capital in Galor and Weil (2000).

The role of scale effects in these models can be illustrated most effectively by looking at the elegant model of Lee (1988). The three key equations of that model are:

$$Y_t = A_t L_t^{1-\beta} T_t^{\beta}, \quad T_t = 1, \tag{67}$$

$$\frac{\dot{A}_t}{A_t} = \gamma \log L_t, \quad A_0 \text{ given}, \tag{68}$$

$$\frac{\dot{L}_t}{L_t} = \alpha\left(\log \frac{Y_t}{L_t} - \log \bar{y}\right), \quad L_0 \text{ given}. \tag{69}$$

Equation (67) describes a production function that depends on ideas A, labor L, and land T, which is assumed to be in fixed supply and normalized to one. Equation (68) is a Romer-like production function for new ideas. Notice that we have assumed the $\phi = 1$ case so that we can get an analytic solution below, but the nature of the results does not depend on this assumption. Notice also that we assume all labor can produce ideas, and we assume a log form. This makes the model log-linear, which is the second key assumption needed to get a closed-form solution. Finally, Equation (69) is a Malthusian equation describing population growth. If output per person is greater than the subsistence parameter $\bar{y}$, then population grows; if less then population declines.

The model can be solved as follows. First, choose the units of output such that the subsistence term gets normalized to zero, $\log \bar{y} = 0$. Next, let $a \equiv \log A$ and $\ell \equiv \log L$. Then the model reduces to a linear homogeneous system of differential equations:

$$\dot{a}_t = \gamma \ell_t, \tag{70}$$

$$\dot{\ell}_t = \alpha a_t - \alpha\beta\ell_t. \tag{71}$$

It is straightforward to solve this system to find

$$\log \frac{Y_t}{L_t} = \omega_1 e^{\theta_1 t} + \omega_2 e^{\theta_2 t}, \tag{72}$$

where $\theta_1 > 0$ and $\theta_2 < 0$ are the eigenvalues associated with this system, and $\omega_1 > 0$.[30] That is, the solution involves a double exponential: the natural log of output per worker

[30] The differential system can be solved using linear algebra, as in Barro and Sala-i-Martin (1995, p. 480) or, even more intuitively, by writing it as a single second-order differential equation, as in Boyce and DiPrima (1997, pp. 123–125). The values for the constants in Equation (72) are $\theta_1 = (-\alpha\beta + \sqrt{(\alpha\beta)^2 + 4\alpha\gamma})/2$, $\theta_2 = (-\alpha\beta - \sqrt{(\alpha\beta)^2 + 4\alpha\gamma})/2$, $\omega_1 = (\theta_1/\alpha)(\alpha(a_0 - \beta\ell_0) - \theta_2\ell_0)/(\theta_1 - \theta_2)$, $\omega_2 = (\theta_2/\alpha)(\theta_1\ell_0 - \alpha(a_0 - \beta\ell_0))/(\theta_1 - \theta_2)$.

grows exponentially, so that the growth rate of output per worker, $y \equiv Y/L$, itself grows exponentially,

$$\frac{\dot{y}_t}{y_t} = \omega_1 \theta_1 e^{\theta_1 t} + \omega_2 \theta_2 e^{\theta_2 t}. \tag{73}$$

Mathematically, it is this double exponential growth that allows the model to deliver a graph that looks approximately like that in Figure 3.

Analytically, Lee's result is extremely nice. However, the analytic results are obtained only by simplifying the model considerably – perhaps too much. For example, the model generates double exponential growth in population as well. As shown in Kremer (1993), this pattern fits the broad sweep of world history, but it sharply contradicts the demographic transition that has set in over the last century, where population growth rates level off and decline. In addition, the analytic results require the strong assumption that $\phi = 1$.

If one wishes to depart from the log-linear structure of Lee's model, the analysis must be conducted numerically. This is done in Jones (2001), with a more realistic demographic setup and with an idea production function that incorporates $\phi < 1$. The basic insights from Lee (1988) apply over the broad course of history, but the model also predicts a demographic transition and a leveling off of per capita income growth in the 20th century. The model with weak scale effects, then, is able to match the basic facts of income and population growth over both the very long run and the 20th century.

The economic intuition for these results is straightforward. Thousands and thousands of years ago, the population was relatively small and the productivity of the population at producing ideas was relatively low. Per capita consumption, then, stayed around the Malthusian level that kept population constant ($\bar{y}$ in the Lee model above). Suppose it took 1000 years for this population to discover a new idea. With the arrival of the new idea, per capita income and fertility rose, producing a larger population. Diminishing returns associated with a fixed supply of land drove consumption back to its subsistence level, but now the population was larger. Instead of requiring 1000 years to produce a new idea, this larger population produced a new idea sooner, say in 800 years. Continuing along this virtuous circle, growth gradually accelerated. Provided the economic environment is characterized by a sufficiently large degree of increasing returns (to offset the diminishing returns associated with limited land), the acceleration in population growth produces a scale effect that leads to the acceleration of per capita income growth. Eventually, the economy becomes sufficiently rich that a demographic transition sets in, leading population growth and per capita income growth to level out.[31]

[31] It is even possible for the demographic transition to drive population growth rates down to zero, in which cases per capita income growth rates decline as well. There is always growth in this world – even a constant population produces new ideas – but the growth rate is no longer exponential. See Jones (2001).

5.5. Summary: scale effects

Virtually all idea-based growth models involve some kind of scale effect, for the basic reason laid out earlier in the presentation of the Idea Diagram. The strong scale effects of many first-generation idea-based growth models – in which the growth rate of the economy is an increasing function of its size – are inconsistent with the relative stability of growth rates in the United States in the 20th century. Subsequent idea-based growth models replaced this strong scale effect with a weak scale effect, where the long-run level of per capita income is an increasing function of the size of the economy. The long-run growth rate in these models is generally an increasing function of the rate of growth of research effort, which in turn depends on the population growth rate of the countries contributing to world research. However, this growth rate is typically taken to be exogenous, producing the policy-invariance results common in these models.

Simple correlations (say of income per person with population, or growth rates of per capita income with population growth rates) on first glance appear to be inconsistent with weak scale effects. However, the ceteris paribus assumption is not valid for such comparisons. Attempts to render other things equal using careful econometrics certainly reveal no inconsistency with the weak scale effects prediction, although they also do not necessarily provide precise estimates of the magnitude of the key scale elasticity.

More broadly, the very long-run history of economic growth appears consistent with weak scale effects. Models in which scale plays an important role have proven capable of explaining the very long-run dynamics of population and per capita income, including the extraordinarily slow growth over much of history and the transition to modern economic growth since the Industrial Revolution.

6. Growth accounting, the linearity critique, and other contributions

This section summarizes a variety of additional insights related to idea-based growth models. Section 6.1 discusses growth accounting in such models, showing that scale effects have accounted for only about 20 percent of U.S. growth in the post-war period. Increases in educational attainment and increases in R&D intensity account for the remaining 80 percent. Section 6.2 considers a somewhat controversial "linearity critique" of endogenous growth models that first appeared in the 1960s. Finally, Section 6.3 will discuss briefly several other important contributions to the literature on growth and ideas that have not yet been mentioned.

6.1. Growth accounting in idea-based models

Growth accounting in a neoclassical framework has a long, illustrious tradition, beginning with Solow (1957). As is well known, such accounting typically finds a residual, which is labeled "total factor productivity growth" (TFP growth). In some ways, endogenous growth models can be understood as trying to find ways to endogenize TFP growth, i.e. to make it something that is determined within the model rather than assumed to be completely exogenous. Having such a model in hand, then, it is quite natural to ask how the model decomposes growth into its sources. That is, quantitatively, how does a particular model account for growth?

Jones (2002b) conducts one of these growth accounting exercises in an economic environment that is basically identical to that analyzed in Section 4. In the long run in that model, per capita growth is proportional to the rate of population growth of the idea-producing regions. Off a balanced growth path, of course, growth can come from transition dynamics, for example, due to capital deepening or to rapid growth in the stock of ideas. Given the stylized fact that U.S. growth rates have been relatively stable over a long period of time, one might be tempted to think that the U.S. is close to its balanced growth path so that growth due to transition dynamics is negligible. On the contrary, however, Jones shows that just the opposite is true. Approximately 80 percent of U.S. growth in the post-war period is due to transition dynamics associated in roughly equal parts with increases in educational attainment and with increases in world R&D intensity. Only about 20 percent of U.S. growth is attributed to the scale effect associated with population growth in the idea-generating countries.[32]

This finding raises a couple of important questions. First, how it is that growth rates can be relatively stable in the United States if transition dynamics are so important? The answer proposed by Jones (2002b) can be seen in a simple analogy. Consider a standard Solow (1956) model that begins in steady state. Now suppose the investment rate increases permanently by 1 percentage point. We know that growth rates rise temporarily and then decline. Now suppose the investment rate, rather than staying constant, grows exponentially. We know that this cannot happen forever since the investment rate is bounded below one. However, it could happen for awhile. In such a world, it is possible for the continued increases in the investment rate to sustain a constant growth rate that is higher than the long-run growth rate. In the idea-based growth model analyzed by Jones (2002b), it is not the investment rate in physical capital that is driving the transition dynamics. Instead, educational attainment and research intensity (the fraction of the labor force working to produce ideas in advanced countries) appear to be rising smoothly in a way that can generate stable growth, at least as an approximation.

The second natural question raised by this accounting concerns the future of U.S. growth. If 80 percent of U.S. growth is due to transition dynamics, then a straightforward implication of the result is that growth rates could slow substantially at some point in the future when the U.S. transits to its balanced growth path. To the extent that population growth rates in the idea-producing countries are declining, this finding is reinforced. Still, there are many other qualifications that must be made concerning this result. Most importantly, it is not clear when the transition dynamics will "run out", particularly since the fraction of the labor force engaged in research seems to be relatively

[32] Comin (2002) suggests that the contribution of R&D to growth could be even smaller. The key assumptions he needs to get this result are that R&D as a share of GDP is truly small, as measured, and that the elasticity of output with respect to ideas is small.

small. In addition, the increased development of countries like China and India means that the pool of potential idea creators could rise for a long time.

6.2. The linearity critique

This section considers the somewhat controversial "linearity critique" of endogenous growth models that first appeared in the 1960s. A coarse version of the criticism is that such models rely on a knife-edge assumption that a particular differential equation is linear in some sense. If the linearity is relaxed slightly, the model either does not generate long-run growth or exhibits growth rates that explode. This section first presents the basic issue and then attempts to show how it can be used productively to make progress in our understanding of economic growth.[33]

Growth models that are capable of producing steady-state growth require strong assumptions. For example, it is well known that steady-state growth is possible only if technological change is labor-augmenting or if the production function is Cobb–Douglas.[34] Another requirement is that the model must possess a differential equation that is linear. That is, all growth models that exhibit steady-state growth ultimately rest on an assumption that some differential equation takes the form

$$\dot{X} = \_\_ X. \tag{74}$$

Growth models differ primarily according to the way in which they label the X variable and the way in which they fill in the blank in this differential equation.[35]

For example, in the Solow (1956) model without technological progress, the differential equation for capital accumulation is less than linear, and the model cannot produce sustained exponential growth. On the other hand, when one adds exogenous technological change in the form of a linear differential equation $\dot{A}_t = gA_t$, one obtains a model with steady-state growth. In the AK growth models of Frankel (1962) and Rebelo (1991), the law of motion for physical capital is assumed to be linear. In the human capital model of Lucas (1988), it is the law of motion for human capital accumulation that is assumed to be linear. Finally, in the first-generation idea-based growth models of Romer (1990), Aghion and Howitt (1992) and Grossman and Helpman (1991), it is the idea production function itself that is assumed to be a linear differential equation.

This kind of knife-edge requirement has made economists uncomfortable for some time. Stiglitz (1990) and Cannon (2000) note that this is one reason endogenous growth models did not catch on in the 1960s even though several were developed.[36] Solow (1994) resurrects this criticism in arguing against recent models of endogenous growth.

[33] This section draws heavily on Jones (2003).
[34] See, for example, the Appendix to Chapter 2 in Barro and Sala-i-Martin (1995).
[35] This approach to characterizing growth models is taken from Romer (1995). Two qualifications apply. First, this linearity can be hidden in models with multiple state variables, as discussed in Mulligan and Sala-i-Martin (1993). Second, linearity is an asymptotic requirement, not a condition that needs to hold at every point in time, as noted by Jones and Manuelli (1990).
[36] The very nice AK model of Frankel (1962) is perhaps the clearest example.

What is not sufficiently well appreciated, however, is that *any* model of sustained exponential growth requires such a knife-edge condition. Neoclassical growth models are not immune to this criticism; they just assume the linearity to be completely unmotivated. One can then proceed in two possible directions. First, one can give up on the desire that a model exhibit steady-state growth. It is not clear where this direction leads, however. One still wants a model to be able to match the steady exponential growth exhibited in the United States for the last 125 years, and it seems likely that a model that produces this kind of behavior will require a differential equation that is nearly linear. Alternatively, one can see the linearity critique as an opportunity for helping us improve our growth models. That is, if a growth model requires a linear differential equation, one can look for an economic explanation for why linearity should hold and/or seek empirical evidence supporting the linearity.

To see how this might work, consider briefly the main types of endogenous growth models and the key differential equations of those models:

1. AK model, $\dot{K} = sK^{\phi}$.
2. Lucas model, $\dot{h} = uh^{\phi}$.
3. R/AH/GH model, $\dot{A} = H_A A^{\phi}$.
4. Fertility model, $\dot{N} = (b - d)N^{\phi}$.

In each case, we can ask the question: “Why should we believe that $\phi \approx 1$ is valid in this model?”. In particular, we consider the following experiment. Suppose we hold constant the individual decision variables (e.g. the investment rate in physical capital or time spent accumulating human capital). Suppose we then double the state variable. Do we double the change in the state variable?

In the AK model, ϕ is the elasticity of output with respect to capital. In the absence of externalities, this elasticity is the share of capital in income. Narrowly interpreting the model as applying to physical capital, one gets a benchmark value of about $1/3$. Some people prefer to include human capital as well, which can get the share a little higher.[37] But then one must appeal to large externalities, and these externalities must be exactly the right size in order to get $\phi \approx 1$.

Now turn to the human capital model of Lucas (1988). Consider a representative agent who lives forever and spends 10 hours per week studying to obtain skills. Are the skills that are added by one period of this studying doubled if the individual’s stock of human capital is doubled? A natural benchmark might be that studying for 10 hours a week adds the same amount, whether one is highly skilled or has little skill. It is far from obvious that the 10 hours of studying increases skills proportionately over time.[38]

[37] I personally think this is a mistake. Human capital is different from physical capital in many ways and gets treated differently in models that are careful about the distinction, e.g. Bils and Klenow (2000).

[38] A subtlety in thinking through the human capital model comes from the Mincerian wage regression evidence. Each year of schooling appears to raise a worker’s wage – and hence productivity – by a constant percentage. One might be tempted to use this to argue that $\phi = 1$ in the human capital case. Bils and Klenow (2000) suggest instead that a human capital accumulation equation of the form $\dot{h} = e^{\theta u} h^{\phi}$ is the right way to capture this evidence.

This chapter has already discussed the Romer/AH/GH assumption of $\phi = 1$. Recall that one can make a case for $\phi < 0$ if it gets harder over time to find new ideas or $\phi > 0$ if knowledge spillovers increase research productivity, or even $\phi = 0$ if researchers produce a constant number of ideas with each unit of effort. The case of $\phi = 1$ appears to have little in the way of intuition or evidence to recommend it.

Finally, the last case above suggests placing the linearity in the equation for population growth, as was done implicitly in the models discussed earlier in the chapter. It can be thought of in this way: Let b and d denote the birth rate and the mortality rate for an individual, respectively. Hold constant an individual's fertility behavior, and suppose we double the number of people in the population. A natural benchmark assumption is that we double the number of offspring. This is the intuition for why a linear differential equation makes sense as a benchmark for the population growth equation.

More generally, I would make the claim that population growth is the least objectionable place to locate a linear differential equation in a growth model, for two reasons. First, if we take population as exogenous and feed in the observed population growth rates into an idea-based growth model, we can explain sustained exponential growth. No additional linearity is needed. Second, the intuition above suggests that it is not crazy to think this differential equation might be close to linear: people reproduce in proportion to their number.[39]

This is one example of how the linearity critique can be used productively. Proponents of particular endogenous growth models can seek evidence and economic insights supporting the hypothesis that the particular engine of growth in a model does indeed involve a differential equation that is close to linear.

6.3. Other contributions

There are a number of other very interesting papers that I have not had time to discuss. These should be given more attention than simply the brief mention that follows, but this chapter is already too long.

Kortum (1997) and Segerstrom (1998) are two important papers that present growth models that exhibit weak scale effects. Both are motivated in part by the stylized fact that total U.S. patents granted to U.S. inventors does not show a large time trend for nearly a century, from roughly 1910 until 1990. If patents are a measure of useful ideas, this fact suggests that the number of new ideas per year might have been relatively stable during a time when per capita income was growing at a relatively constant rate. How can this be? In the models provided above, the stock of ideas grows at a constant rate, just like output. Kortum (1997) and Segerstrom (1998) solve this puzzle by supposing that ideas, at least on average, represent *proportional* improvements in productivity. The

[39] This does not mean that fertility behavior, b, will ensure a positive rate of population growth forever. That is a different question. Indeed, Jones (2001) supposes that a demographic transition ultimately leads to zero population growth in attempting to explain growth over the very long run.

papers also assume that new ideas are increasingly difficult to obtain, so that in steady state, a growing number of researchers produce a constant number of new ideas, which in turn leads to a constant rate of exponential growth.

Romer (1994a) makes an interesting point that appears (at least based on citations) to have been under appreciated in the literature. The paper considers the welfare cost of trade restrictions from the standpoint of models in which ideas play an important role. In neoclassical models, trade restrictions, like other taxes, typically have small effects associated with Harberger triangles that depend on the square of the tax rate. In contrast, Romer shows that if trade restrictions reduce the range of goods (ideas) available within a country, the welfare affect is proportional to the level of the tax rate rather than its square. As a result, distortions that affect the use of ideas can have much larger welfare effects than those same distortions in neoclassical models.

Acemoglu (2002) surveys a number of important results that come from thinking about the direction of technological change. In this general framework, researchers can choose to search for ideas that augment different factors. For example, they may search for ideas that augment capital or skilled labor or unskilled labor. Other things equal, a market size effect suggests that research will be targeted toward augmenting factors that are in greater supply, especially when these factors can be easily substituted for other factors of production.

Greenwood, Hercowitz and Krusell (1997) and Whelan (2001) focus on the rapid technological change that is associated with the declines in the relative prices of consumer and producer durables (driven in large part by the rapid declines in the quality-adjusted price of semiconductors). Greenwood, Hercowitz and Krusell (1997) show that investment-specific technological change can account for roughly half of per capita income growth in the United States in recent decades. Whelan (2001) extends this analysis by tying it to the introduction of chained indexes in the national income and product accounts.

Finally, it is worth mentioning again that this chapter has largely omitted a very important part of the literature on growth and ideas, that associated with the Schumpeterian models of Aghion and Howitt (1992) and Grossman and Helpman (1991). These models were applied in detail to international trade in Grossman and Helpman (1991). Aghion and Howitt (1998) contain a rich analysis of an even wider range of applications, to such topics as unemployment, the effects of increases in competition, patent races, and leader-follower effects in R&D. In addition to these excellent treatments, a separate Handbook chapter by Aghion and Howitt surveys some of these important topics.

7. Conclusions

Thinking carefully about the way in which ideas are different from other economic goods leads to a profound change in the way we understand economic growth. The nonrivalry of ideas implies that increasing returns to scale is likely to characterize production possibilities. This leads to a world in which scale itself can serve as a source of long run growth. The more inventors we have, the more ideas we discover, and the richer we all are. This also leads to a world where the first fundamental welfare theorem no long necessarily holds. Perfectly competitive markets may not lead to the optimal allocation of resources. This means that other institutions may be needed to improve welfare. The patent system and research universities are examples of such institutions, but there is little reason to think we have found the best institutions – after all these institutions are themselves ideas.

While we have made much progress in understanding economic growth in a world where ideas are important, there remain many open, interesting research questions. The first is "What is the shape of the idea production function?" How do ideas get produced? The combinatorial calculations of Romer (1993) and Weitzman (1998) are fascinating and suggestive. The current research practice of modeling the idea production function as a stable Cobb–Douglas combination of research and the existing stock of ideas is elegant, but at this point we have little reason to believe that it is correct. One insight that illustrates the incompleteness of our knowledge is that there is no reason why research productivity in the idea production function should be a smooth, monotonic function of the stock of ideas. One can easily imagine that some ideas lead to a domino-like unraveling of phenomena that were previously mysterious, much like the general purpose technologies of Helpman (1998). Indeed, perhaps the decoding of the human genome or the continued boom in information technology will lead to a large upward shift in the production function for ideas.[40] On the other hand, one can equally imagine situations where research productivity unexpectedly stagnates, if not forever then at least for a long time. Progress in the time it takes to travel from New York to San Francisco represents a good example of this.

A second unresolved research question is "What is the long-run elasticity of output per worker with respect to population?". That is, how large are increasing returns to scale. This parameter (labeled γ in the main models of this chapter) is crucially related to the long-run rate of growth of the economy. Estimating it precisely would not only provide confirmation of idea-based growth theory but would also help us in accounting for the sources of economic growth.

Finally, a policy-related question: "What are better institutions and policies for encouraging the efficient amount of research?". There is a large, suggestive literature on social rates of return to research and on the extent to which firms might underinvest in research. Still, none of these individual studies is especially compelling, and more accurate estimates of these gaps would be valuable. To the extent that the returns to research do not reflect the marginal benefit to society, better institutions might improve allocations.

40 Dale Jorgenson, in his Handbook chapter, suggests that the information technology revolution may do just this.

Acknowledgements

I am grateful to Philippe Aghion and Paul Romer for helpful comments, and to the National Science Foundation for research support.

References

Acemoglu, D. (2002). "Directed technical change". Review of Economic Studies 69 (4), 781–810.

Aghion, P., Howitt, P. (1992). "A model of growth through creative destruction". Econometrica 60 (2), 323–351.

Aghion, P., Howitt, P. (1998). Endogenous Growth Theory. MIT Press, Cambridge, MA.

Alcala, F., Ciccone, A. (2002). "Trade and productivity". Mimeo. Universitat Pompeu Fabra.

Arrow, K.J. (1962a). "The economic implications of learning by doing". Review of Economic Studies 29 (June), 153–173.

Arrow, K.J. (1962b). "Economic welfare and the allocation of resources for invention". In: Nelson, R.R. (Ed.), The Rate and Direction of Inventive Activity. Princeton University Press (NBER), Princeton, NJ, pp. 609–625.

Atkeson, A.A., Kehoe, P.J. (2002). "The transition to a new economy following the second industrial revolution". NBER Working Paper 8676.

Backus, D.K., Kehoe, P.J., Kehoe, T.J. (1992). "In search of scale effects in trade and growth". Journal of Economic Theory 58 (August), 377–409.

Barro, R.J., Sala-i-Martin, X. (1995). Economic Growth. McGraw-Hill, New York.

Ben-David, D., Papell, D.H. (1995). "The great wars, the great crash, and steady-state growth: Some new evidence about an old stylized fact". Journal of Monetary Economics 36 (December), 453–475.

Bils, M., Klenow, P. (2000). "Does schooling cause growth?". American Economic Review 90 (December), 1160–1183.

Boldrin, M., Levine, D.K. (2002). "Perfectly competitive innovation". FRB of Minneapolis Staff Report 303.

Boserup, E. (1965). The Conditions of Agricultural Progress. Aldine Publishing Company, Chicago, IL.

Boyce, W.E., DiPrima, R.C. (1997). Elementary Differential Equations and Boundary Value Problems. Wiley, New York.

Cannon, E. (2000). "Economies of scale and constant returns to capital: A neglected early contribution to the theory of economic growth". American Economic Review 90 (1), 292–295.

Clark, G. (2001). "The secret history of the industrial revolution". Mimeo. University of California, Davis.

Comin, D. (2002). "R&D? A small contribution to productivity growth". Mimeo. New York University.

Cozzi, G. (1997). "Exploring growth trajectories". Journal of Economic Growth 2 (4), 385–399.

David, P.A. (1993). "Knowledge, property, and the system dynamics of technological change". Proceedings of the World Bank Annual Conference on Development Economics, 1992, 215–248.

DeLong, J.B. (2000). "Slouching Toward Utopia". University of California, Berkeley. Online at: http://econ161.berkeley.edu/TCEH/Slouch_Old.html.

Diamond, J. (1997). Guns, Germs, and Steel. W.W. Norton and Co., New York.

Dinopoulos, E., Thompson, P. (1998). "Schumpeterian growth without scale effects". Journal of Economic Growth 3 (4), 313–335.

Dixit, A.K., Stiglitz, J.E. (1977). "Monopolistic competition and optimum product diversity". American Economic Review 67 (June), 297–308.

Ethier, W.J. (1982). "National and international returns to scale in the modern theory of international trade". American Economic Review 72 (June), 389–405.

Frankel, J.A., Romer, D. (1999). "Does trade cause growth?". American Economic Review 89 (3), 379–399.

Frankel, M. (1962). "The production function in allocation and growth: A synthesis". American Economic Review 52 (December), 995–1022.

Galor, O., Weil, D. (2000). "Population, technology, and growth: From the Malthusian regime to the demographic transition". American Economic Review 90 (September), 806–828.

Goodfriend, M., McDermott, J. (1995). "Early development". American Economic Review 85 (1), 116–133.

Greenwood, J., Hercowitz, Z., Krusell, P. (1997). "Long-run implications of investment-specific technological change". American Economic Review 87 (3), 342–362.

Greenwood, J., Seshadri, A., Yorukoglu, M. (2001). "Engines of liberation". Mimeo. University of Rochester.

Griliches, Z. (1998). R&D and Productivity: The Econometric Evidence. University of Chicago Press.

Grossman, G.M., Helpman, E. (1991). Innovation and Growth in the Global Economy. MIT Press, Cambridge, MA.

Hall, R.E., Jones, C.I. (1999). "Why do some countries produce so much more output per worker than others?". Quarterly Journal of Economics 114 (1), 83–116.

Hansen, G.D., Prescott, E.C. (2002). "Malthus to Solow". American Economic Review 92 (4), 1205–1217.

Helpman, E. (Ed.) (1998). General Purpose Technologies and Economic Growth. MIT Press, Cambridge, MA.

Howitt, P. (1999). "Steady endogenous growth with population and R&D inputs growing". Journal of Political Economy 107 (4), 715–730.

Jones, C.I. (1995a). "R&D-based models of economic growth". Journal of Political Economy 103 (4), 759–784.

Jones, C.I. (1995b). "Time series tests of endogenous growth models". Quarterly Journal of Economics 110 (441), 495–525.

Jones, C.I. (1999). "Growth: With or without scale effects?". American Economic Association Papers and Proceedings 89 (May), 139–144.

Jones, C.I. (2001). "Was an Industrial Revolution inevitable? Economic growth over the very long run". Advances in Macroeconomics 1 (2). Article 1. http://www.bepress.com/bejm/advances/voll/iss2/artl.

Jones, C.I. (2002a). Introduction to Economic Growth, second ed. W.W. Norton and Co., New York.

Jones, C.I. (2002b). "Sources of U.S. economic growth in a world of ideas". American Economic Review 92 (1), 220–239.

Jones, C.I. (2003). "Population and ideas: A theory of endogenous growth". In: Aghion, P., Frydman, R., Stiglitz, J., Woodford, M. (Eds.), Knowledge Information and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps. Princeton University Press, Princeton, NJ, pp. 498–521.

Jones, C.I., Williams, J.C. (1998). "Measuring the social return to R&D". Quarterly Journal of Economics 113 (4), 1119–1135.

Jones, L.E., Manuelli, R. (1990). "A convex model of economic growth: Theory and policy implications". Journal of Political Economy 98 (October), 1008–1038.

Judd, K.L. (1985). "On the performance of patents". Econometrica 53 (3), 567–585.

Kortum, S.S. (1997). "Research, patenting, and technological change". Econometrica 65 (6), 1389–1419.

Kremer, M. (1993). "Population growth and technological change: One million B.C. to 1990". Quarterly Journal of Economics 108 (4), 681–716.

Kremer, M. (1998). "Patent buyouts: A mechanism for encouraging innovation". Quarterly Journal of Economics 113 (November), 1137–1167.

Kuznets, S. (1960). "Population change and aggregate output". In: Demographic and Economic Change in Developed Countries. Princeton University Press, Princeton, NJ.

Lee, R.D. (1988). "Induced population growth and induced technological progress: Their interaction in the accelerating stage". Mathematical Population Studies 1 (3), 265–288.

Li, C.-W. (2000). "Endogenous vs. semi-endogenous growth in a two-R&D-sector model". Economic Journal 110 (462), C109–C122.

Li, C.-W. (2002). "Growth and scale effects: The role of knowledge spillovers". Economics Letters 74 (2), 177–185.

Lipscomb, A.A., Bergh, A.E. (1905). The Writings of Thomas Jefferson, vol. 13. Thomas Jefferson Memorial Association, Washington. Online at http://press-pubs.uchicago.edu/founders/documents/al_8_8sl2.html.

Lucas, R.E. (1988). "On the mechanics of economic development". Journal of Monetary Economics 22 (1), 3–42.

Lucas, R.E. (1998). “The Industrial Revolution: Past and Future”. Mimeo. University of Chicago.
Maddison, A. (1995). Monitoring the World Economy 1820–1992. Organization for Economic Cooperation and Development, Paris.
Mankiw, N.G., Romer, D., Weil, D. (1992). “A contribution to the empirics of economic growth”. Quarterly Journal of Economics 107 (2), 407–438.
Mulligan, C.B., Sala-i-Martin, X. (1993). “Transitional dynamics in two-sector models of endogenous growth”. Quarterly Journal of Economics 108 (3), 739–774.
Nordhaus, W.D. (1969). “An economic theory of technological change”. American Economic Association Papers and Proceedings 59 (May), 18–28.
Nordhaus, W.D. (2003). “The health of nations: The contribution of improved health to living standards”. In: Murphy, K.M., Topel, R. (Eds.), Measuring the Gains from Medical Research: An Economic Approach. University of Chicago Press, Chicago, IL, pp. 9–40.
Ohanian, L., Cole, H. (2001). “Re-examining the contribution of money and banking shocks to the U.S. great depression”. NBER Macroeconomics Annual 16, 183–227.
Peretto, P. (1998). “Technological change and population growth”. Journal of Economic Growth 3 (4), 283–311.
Phelps, E.S. (1966). “Models of technical progress and the golden rule of research”. Review of Economic Studies 33 (2), 133–145.
Phelps, E.S. (1968). “Population increase”. Canadian Journal of Economics 1 (3), 497–518.
Quah, D.T. (1996). “The Invisible Hand and the Weightless Economy”. Mimeo. LSE Economics Department. April.
Quah, D. (2002). “Almost efficient innovation by pricing ideas”. Mimeo. LSE. June.
Rebelo, S. (1991). “Long-run policy analysis and long-run growth”. Journal of Political Economy 99 (June), 500–521.
Rivera-Batiz, L.A., Romer, P.M. (1991). “Economic integration and endogenous growth”. Quarterly Journal of Economics 106 (May), 531–555.
Romer, P.M. (1986). “Increasing returns and long-run growth”. Journal of Political Economy 94 (October), 1002–1037.
Romer, P.M. (1987). “Growth based on increasing returns to specialization”. American Economic Review Papers and Proceedings 77 (May), 56–62.
Romer, P.M. (1990). “Endogenous technological change”. Journal of Political Economy 98 (5), S71–S102.
Romer, P.M. (1993). “Two strategies for economic development: Using ideas and producing ideas”. Proceedings of the World Bank Annual Conference on Development Economics 1992, 63–115.
Romer, P.M. (1994a). “New goods, old theory, and the welfare costs of trade restrictions”. Journal of Development Economics 43, 5–38.
Romer, P.M. (1994b). “The origins of endogenous growth”. Journal of Economic Perspectives 8, 3–22.
Romer, P.M. (1995). “Comment on a paper by T.N. Srinivasan”. In: Growth Theories in Light of the East Asian Experience. University of Chicago Press, Chicago, IL.
Romer, P.M. (2000). “Should the government subsidize supply or demand in the market for scientists and engineers?”. NBER Working Paper 7723, June.
Sala-i-Martin, X. (1997). “I just ran four million regressions”. NBER Working Paper 6252.
Segerstrom, P. (1998). “Endogenous growth without scale effects”. American Economic Review 88 (5), 1290–1310.
Segerstrom, P.S., Anant, T.C.A., Dinopoulos, E. (1990). “A Schumpeterian model of the product life cycle”. American Economic Review 80 (5), 1077–1091.
Shell, K. (1966). “Toward a theory of inventive activity and capital accumulation”. American Economic Association Papers and Proceedings 56, 62–68.
Sheshinski, E. (1967). “Optimal accumulation with learning by doing”. In: Shell, K. (Ed.), Essays on the Theory of Economic Growth. MIT Press, Cambridge, MA, pp. 31–52.
Simon, J.L. (1986). Theory of Population and Economic Growth. Blackwell, New York.
Simon, J.L. (1998). The Ultimate Resource 2. Princeton University Press, Princeton, NJ.

Solow, R.M. (1956). "A contribution to the theory of economic growth". Quarterly Journal of Economics 70 (1), 65–94.

Solow, R.M. (1957). "Technical change and the aggregate production function". Review of Economics and Statistics 39 (3), 312–320.

Solow, R.M. (1994). "Perspectives on growth theory". Journal of Economic Perspectives 8 (1), 45–54.

Spence, M. (1976). "Product selection, fixed costs, and monopolistic competition". Review of Economic Studies 43 (2), 217–235.

Stiglitz, J.E. (1990). "Comments: Some retrospective views on growth theory". In: Diamond, P. (Ed.), Growth/Productivity/Unemployment: Essays to Celebrate Robert Solow's Birthday. MIT Press, Cambridge, MA, pp. 50–69.

Stokey, N.L. (2001). "A quantitative model of the British industrial revolution, 1780–1850". Carnegie Rochester Conference Series on Public Policy 55, 55–109.

Tamura, R. (2002). "Human capital and the switch from agriculture to industry". Journal of Economic Dynamics and Control 27, 207–242.

Weitzman, M.L. (1976). "On the welfare significance of national product in a dynamic economy". Quarterly Journal of Economics 90 (1), 156–162.

Weitzman, M.L. (1998). "Recombinant growth". Quarterly Journal of Economics 113 (May), 331–360.

Whelan, K. (2001). "Balance growth revisited: A two-sector model of economic growth". Mimeo. Federal Reserve Board of Governors.

Williams, J.C. (1995). "Tax and Technology Policy in an Endogenous Growth Model". Ph.D. Dissertation. Stanford University.

Young, A. (1998). "Growth without scale effects". Journal of Political Economy 106 (1), 41–63.