

NBER WORKING PAPER SERIES

ENDOGENOUS TECHNOLOGICAL CHANGE

Paul M. Romer

Working Paper No. 3210

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 1989

prepared for the conference "The Problem of Economic Development," SUNY Buffalo, May 1988. I have benefitted from the comments of many seminar and conference participants and two discussants (Rob Vishny, Buffalo May 1988, and Dale Jorgenson, NBER Economic Fluctuations meeting, July, 1988). Discussions with Gary Becker, Karl Shell, Robert Lucas, Gene Grossman, and Elhanan Helpman were especially helpful. Research assistance was provided by Danyang Xie. The original work was supported by NSF grant #SES-8618325. It was revised while I was a visitor at the Center for Advanced Study in the Behavioral Sciences and supported by NSF grant #BNS87-00864. This paper is part of NBER's research program in Growth. Any opinions expressed are those of the author not those of the National Bureau of Economic Research.

NBER Working Paper #3210
December 1989

ENDOGENOUS TECHNOLOGICAL CHANGE

ABSTRACT

Growth in this model is driven by technological change that arises from intentional investment decisions made by profit maximizing agents. The distinguishing feature of the technology as an input is that it is neither a conventional good nor a public good; it is a nonrival, partially excludable good. Because of the nonconvexity introduced by a nonrival good, price-taking competition cannot be supported, and instead, the equilibrium is one with monopolistic competition. The main conclusions are that the stock of human capital determines the rate of growth, that too little human capital is devoted to research in equilibrium, that integration into world markets will increase growth rates, and that having a large population is not sufficient to generate growth.

Paul M. Romer
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637

1. Introduction

Output per hour worked in the United States today is ten times as valuable as output per hour worked 100 years ago. (Maddison, 1982). Since the 1950s, economists have attributed much of the change in output per hour worked either directly or indirectly to technological change. (Abromowitz, 1956, Kendrick, 1956, Solow, 1957.) From a naive point of view this seems right. The raw materials that we use have not changed, but as a result of trial and error, experimentation, refinement, and scientific investigation, the instructions that we follow for combining raw materials have become vastly more sophisticated. One hundred years ago, all we could do to get visual stimulation from iron oxide was to use it as pigment. Now we put it on plastic tape and use it to make video cassette recordings.

The argument presented in this paper is based on three premises. The first is that technological change—improvement in the instructions for mixing together raw materials—lies at the heart of economic growth. As a result, the model presented here resembles the Solow (1956) model with technological change. Technological change provides the incentive for continued capital accumulation, and together, capital accumulation and technological change account for much of the increase in output per hour worked.

The second premise is that technological change arises in large part because of intentional actions taken by people who respond to market incentives. Thus, the model is one of endogenous rather than exogenous technological change. This does not mean that everyone who contributes to technological change is motivated by market incentives. An academic scientist who is supported by government grants may be totally insulated from them. The premise here is that market incentives nonetheless play an essential role in the process whereby new knowledge is translated into goods with practical value. Our initial

understanding of electromagnetism arose from research conducted in academic institutions, but magnetic tape and home video cassette recorders resulted from attempts by private firms to earn a profit.

The third and most fundamental premise is that instructions for working with raw materials are inherently different from other economic goods. Once the cost of creating a new set of instructions has been incurred, the instructions can be used over and over again at no additional cost. Developing new and better instructions is equivalent to incurring a fixed cost. This property is taken to be the defining characteristic of technology.

Most models of aggregate growth, even those with spillovers or external effects, rely on price-taking behavior. But once these three premises are granted, it follows directly that an equilibrium with price-taking cannot be supported. Section 2 of the paper starts by showing why this is so. It also indicates which of the the premises is dropped in growth models that do depend on price-taking behavior. The argument in this section is fundamental to the motivation for the particular model of monopolistic competition that follows, but is more general than the model itself.

In the specific model outline in Section 3, a firm incurs fixed design or research and development costs when it creates a new good. It recovers those costs by selling the new good for a price that is higher than its constant cost of production. Since there is free entry into this activity, firms earn zero profit in a present value sense.

The conclusions of the model follow directly from this specification. Based on results from the static theory of trade with differentiated goods (see for example Helpman and Krugman 1985), one should expect that fixed costs lead to gains from increases in the size of the market, and therefore to gains from trade between different countries. Perhaps the most interesting feature of the equilibrium calculated for the model constructed here is that increases in the size of the market have effects not only on the level of income and welfare, but also on the rate of growth. Larger markets induce more research and faster growth.

The analysis also suggests why population is not the right measure of market size, and why the presence of a large domestic market in countries like China or India is not a substitute for trade with the rest of the world. The growth rate is increasing in the stock of human capital, but it does not depend on the total size of the labor force or the population. In a limiting case that may have been relevant for historical analysis and for the poorest countries today, if the stock of human capital is too low, growth may not take place at all.

These implications of the model are taken up briefly in the final sections of the paper. Section 3 describes the functional forms that are used to describe the preferences and the technology for the model. It defines an equilibrium that allows for both monopolistic competition and external effects arising from knowledge spillovers. Section 4 offers a brief intuitive description of a balanced growth equilibrium for the model. Section 5 formally characterizes the equilibrium and compares it to the social optimum that could be achieved by an omniscient, omnipotent planner. The paper closes with a discussion of the implications of the model concerning trade, research, and growth. Algebraic details of the derivations are placed in an appendix.

2. Rivalry, excludability, and nonconvexities

Economists studying public finance have identified two fundamental attributes of any economic good: the degree to which it is rivalrous and the degree to which it is excludable. (Cornes and Sandler, 1986) Rivalry is a purely technological attribute. A purely rival good has the property that its use by one firm or person precludes its use by another; a purely nonrival good has the property that its use by one firm or person in no way limits its use by another. Excludability is a function of both the technology and the legal system. A good is excludable if the owner of the good can prevent others from using

it. A good like the code for a computer program can be made excludable by means of a legal system that sanctions copying or by means of encryption and copy protection schemes.

Conventional economic goods are both rivalrous and excludable. They are privately provided and can be traded in competitive markets. By definition, public goods are both nonrival and nonexcludable. Because they are nonexcludable, they cannot be privately provided or traded in markets. Public goods can be introduced into a model of price-taking behavior by assuming the existence of a government that can levy taxes. Basic scientific research is an example of a public good that could be provided in this way and that is relevant for modeling growth.

Rivalry and excludability are closely linked because there is no interesting sense in which a good can be rival but nonexcludable. For a rival good, the ability to use the good necessarily implies the ability to exclude others from doing so. Problems of spillovers or incomplete appropriability do not arise.

A good can, however, be excludable and nonrival. The third premise cited in the introduction implies that technology is a nonrival input. The second premise implies that technological change takes place because of the actions of self-interested individuals, so improvements in the technology must confer benefits that are at least partially excludable. The first premise therefore implies that growth is driven fundamentally by the accumulation of a partially excludable, nonrival input.

To evaluate these claims, it helps to have a specific case in mind. The example of a nonrival inputs used in what follows is a design for a new good. The vast majority of designs result from the research and development activities of private, profit maximizing firms. A design is, nonetheless, nonrival. Once the design is created, it can be used as often as desired, in as many productive activities as desired.

In this sense, a design differs in a crucial way from a piece of human capital like the ability to add. The design is nonrival but the ability to add is not. The difference arises

because the ability to add is inherently tied to a physical object (a human body) whereas the design is not.¹ The ability to add is rivalrous because the person who possesses this ability cannot be in more than one place at the same time; nor can this person solve many problems at once. By the argument noted above, rivalry implies that human capital is also excludable. Thus, human capital can be privately provided and traded in competitive markets. In contrast, the design is nonrival because it is independent of any physical object. It can be copied and used in as many different activities as desired.

Like any scientific concept, nonrivalry is an idealization. It is sometimes observed that a design cannot be a nonrival good because it is itself tied to the physical piece of paper or the physical computer disk on which it is stored. What is unambiguously true about a design is that the cost of replicating it with a draftsman, a photocopier, or a disk drive is trivial compared to the cost of creating the design in the first place. This is not true of the ability to add. Training the second person to add is as costly as training the first. For simplicity, the arguments here will treat designs as idealized goods that are not tied to any physical good and can be costlessly replicated, but nothing hinges on whether this is literally true or merely close to being true.

Nonrivalry has two important important implications for the theory of growth. First, nonrival goods can be accumulated without bound on a per capita basis, whereas a piece of human capital like the ability to add cannot. Each person has only a finite number of years that can be spent acquiring skills. When this person dies, the skills are lost, but any nonrival good that this person produces—a scientific law, a principle of mechanical, electrical and chemical engineering, a mathematical result, software, a patent, a mechanical drawing, or a blueprint—lives on after the person is gone. Second, treating

¹The original version of this paper used the terms "embodied" and "disembodied" to refer to the difference between an intangible like the ability to add that is tied to a specific person and an intangible like a design that is not. This choice of terminology is not used in this revision because embodiment has another meaning in growth theory and because the notion of rivalry already exists in the public finance literature.

knowledge as a nonrival good makes it possible to talk sensibly about knowledge spillovers, i.e. incomplete excludability. These two features of knowledge—unbounded growth and incomplete appropriability—are features that are generally recognized as being relevant for the theory of growth. What thinking about nonrivalry shows is that these features are inextricably linked to nonconvexities.

If a nonrival input has productive value, then output cannot be a constant returns to scale function of all of its inputs taken together. The standard replication argument used to justify homogeneity of degree one does not apply because it is not necessary to replicate nonrival inputs. Suppose that a firm can invest 10,000 hours of engineering time to produce a design for a 20 megabyte hard disk drive for computers. Suppose it can produce a total of 2 trillion megabytes of storage per year (i.e. 100,000 units of the drive) if it builds a \$10 million factory and hires 100 workers. If it merely replicates the rival inputs—the factory and the workers—it can double its output to 4 trillion megabytes of storage per year.

Suppose that the firm could have invested 20,000 hours of engineering time in the design work instead of 10,000 hours, and by so doing, could have designed a 30 megabyte hard disk drive that could be manufactured with the same factory and workers. When the firm doubles all of its inputs, it uses a 20,000 hour design, 2 factories, and 200 workers and produces 6 trillion megabytes of storage per year, 3 times the output original output.

More formally, if $F(A,X)$ represents a production process that depends on rival inputs X and nonrival inputs A , then by a replication argument, it follows that

$$F(A,\lambda X) = \lambda F(A,X).$$

This replication argument assumes that the list X is an exhaustive list of the rival inputs. Because the focus here is on national economies, the argument neglects integer problems that may be relevant for a small market that gets stuck between n and $n+1$ plants. The

fact that it may not actually be possible to actually replicate all of the inputs in the list X has no bearing on this argument about the properties of $F(\cdot)$.

If A is productive as well, it follows that F cannot be a concave production function because

$$F(\lambda A, \lambda X) > \lambda F(A, X).$$

Because of the properties of homogeneous functions, it also follows that a firm with this kind of production possibilities could not survive as a price-taker. If disk drives sold for marginal cost, annual revenue for the firm would just equal interest payments on the capital and wage payments to workers. More generally, since $F(A, Z) = X \cdot \frac{\partial F}{\partial X}(A, X)$, it follows that

$$F(A, Z) < A \cdot \frac{\partial F}{\partial A}(A, X) + X \cdot \frac{\partial F}{\partial X}(A, X).$$

If all inputs were paid their value marginal product, the firm would suffer losses.

This point has been made many times before. (Schumpeter 1942, Arrow, 1962b, Shell, 1966, 1967, 1973, Nordhaus, 1969, and Wilson 1975.) Previous growth models have avoided this difficulty in various ways. Solow (1956) treats A as an exogenously provided public input (i.e. an input that is both nonexcludable and nonrival). Shell (1966, 1967) treats it as a public input that is provided by the government. In each case, the factor A receives no compensation, and every individual firm is assumed to be free to exploit the entire stock of A . These models are consistent with the first premise, that technological change drives growth, and the third, that the technology is a nonrival good, but they are inconsistent with the second premise. They both deny the role that private, maximizing behavior plays in generating technological change.

In an attempt to make the evolution of A responsive to market incentives, Arrow

(1962a) assumed that an increase in K necessarily leads to an equiproportionate increase in knowledge through "learning-by-doing", but still treats knowledge as a public good. Lucas (1988) assumes in effect that it is production of human capital rather than physical capital that generates this nonrival, nonexcludable good. Both of these papers make the production of a nonrival, nonexcludable good an unintentional side effect of the production of a conventional good.

The learning-by-doing formulation has the advantage that it makes the rate of accumulation of nonrival knowledge endogenous, but it is unsatisfactory because it takes the strict proportionality between knowledge and physical capital or knowledge and education as an unexplained and exogenously given feature of the technology. It preserves the public good character of knowledge assumed by Solow and Shell, but makes it a public good that is privately provided as a side effect. Like the other public good formulations, it rules out the possibility that firms make intentional investments in research and development.

This formulation has the additional difficulty that it is not robust. The nonrival input produced through learning-by-doing must be completely nonexcludable. If it were even partially excludable, Dasgupta and Stiglitz (1988) show that decentralized equilibrium with many firms would not be sustainable.

In a partial equilibrium model of an industry in which firms face upward sloping cost curves, Shell (1973) proposed a model with price-taking in which expenditure on research was compensated out of quasi-rents. Griliches (1979), again in an industry setting, made this formulation more explicit. He assumed that the production function takes the form $F(A_N, A_E, X)$, where A_E represents an excludable part the benefits of research and development and A_N represents the nonexcludable part. Since A_E is excludable, it is accumulated intentionally. The nonexcludable part A_N is created as a side effect of producing A_E . He also assumed that the function $F(\cdot)$ is homogeneous of degree 1 in X and A_E taken together.

In an aggregate model of growth, I made the same kind of assumption (Romer 1986). To make the dynamic analysis in this paper simple, I reduced the dynamic model to one with a single state variable model by assuming that the excludable good A_E that the firm produces intentionally is used in fixed proportions with physical capital. As a result, the model ends up having dynamic similar to those of Arrow's learning-by-doing model and the mathematical equations presented can equally well be interpreted along the lines suggested by Arrow.

The advantage of this type of formulation is that it allows for intentional private investments in research and development. The difficulty is that it violates the logic of the replication argument. If the input A_E is truly the result of research and development, it is a nonrival good. In this case, the function $F(\cdot)$ must be homogeneous of degree 1 in X alone and the Dasgupta and Stiglitz (1987) argument applies. If a nonrival input is even partially excludable, nonconvexities are present and a decentralized equilibrium cannot be sustained. Stated in terms of cost curves, the replication argument implies that long run cost curves are horizontal when all rival inputs are treated as being variable. What appear to be quasi-rents are merely competitive returns to rival factors that are in fixed supply. These quasi-rents cannot be used to compensate both the innovation activity and the rival, fixed factors. Even if these factors are in fixed supply at the aggregate level, they presumably have alternative uses and would not be supplied to an activity if they are not paid their marginal product.

An alternative approach to growth theory with price-taking competition is to dispute the first and third premises, that technology change drives growth and that knowledge about the technology is a nonrival input. Human capital models like presented by King and Rebelo (1987), Rebelo (1988), Becker, Murphy and Tamura (1989), and Jones and Manuelli (1988) treat all forms of intangible knowledge as being analogous to human capital skills that are rivalous and excludable. There is no nonrival input like the technology, hence, no nonconvexities or spillovers.

The only way to accept all three of the premises described in the introduction is to return to the suggestion of Schumpeter (1942) and explicitly introduce market power. Shell (1973) described a model with a single monopolist who invests in technological change, but as he recognizes, it is difficult to give an aggregate interpretation to a model dominated by a single firm. In a recent paper (Romer, 1987) I presented a model with market power but also with free entry and many firms. It builds on the model of monopolistic competition in consumption goods formulated by Dixit and Stiglitz (1972), applied in a dynamic setting by Judd (1985), and extended to differentiated inputs in production by Either (1982). Like the present model, this model has the property that the size of the market affects the rate of growth, but the motivation for the presence of the fixed cost is not well articulated, and a separate role for human capital in the production of technological innovation is not emphasized. The extension of this model presented here tries to remedy these defects.

3. Description of the Model

The four basic inputs in this model are capital, labor, human capital, and an index of the level of the technology. Capital is measured in units of consumption goods. Labor services L are skills like eye-hand coordination that are available from a healthy physical body. They are measured by counts of people. As used here, human capital H is a distinct measure of the cumulative effect of activities like formal education and on-the-job training.

The concept of human capital as years of education or training that are person specific is close to the one used in labor market contexts (e.g., Rosen 1976 and Heckman 1976). It corresponds to the practice in growth accounting applications that take account of changes in the quality of the labor force due to changes in observables like the level of

education and experience. (See for example Gollop and Jorgenson 1980.) This concept of human capital is more limited than the notion used in theoretical models of growth based on unlimited human capital accumulation such as those presented by Lucas (1988), King and Rebelo (1987), and Becker, Murphy, and Tamura (1989). These models implicitly combine a notion of knowledge that can outlive any individual with a the labor market notion of human capital that does not.

The model used here separates the rival component of knowledge, H , from the nonrival, technological component, A . Because it has an existence that is separate from that of any individual, A can grow without bound. In the specific formulation used below, each new unit of knowledge corresponds to a design for a new good, so there is no conceptual problem measuring A . It is a count of the number of designs.

The formal model of the economy has three sectors. The research sector uses human capital and the existing stock of knowledge to produce new knowledge. Specifically, it produces designs for new producer durables. An intermediate goods sector uses the designs from the research sector together with capital to produce the large number of producer durables that are available for use in final goods production at any time. In practice, one might expect research on a new design and the production of the new good to take place within the same firm, and nothing in the analysis here rules this out; design work can take place either internally or in a separate firm that sells its patent to the firm that will produce the actual good. A final goods sector uses labor, human capital, and the set of producer durables that are available to produce final output. Output can be either consumed or saved as new capital.

To keep the dynamic analysis simple and highlight the effects of interest, several simplifying assumptions are used. The first is that the population and the supply of labor are both constant. This rules out an analysis of fertility, labor force participation, or variation in hours worked per worker. The second is that the total stock of human capital in the population is fixed and that the fraction supplied to the market is also fixed. Thus

the supply of the aggregate factors L and H is fixed. The assumption on H is made largely for technical reasons. The dynamic analysis is greatly simplified by restricting attention to equilibria with constant growth rates. In a stationary population where people have finite lives, the only feasible constant growth rate for total years of education or experience is zero. A more complicated dynamic analysis could consider the effects of the kinds of increases in H and L that have been observed historically.

The other simplifying assumptions are extreme assumptions on factor intensities. One has already been made implicitly. Assuming that capital can be accumulated as forgone output is equivalent to assuming that capital goods are produced in a separate sector that has the same technology as the final output sector. Forgoing consumption is then equivalent to shifting resources from the consumption sector into the capital sector. Also, the plausible assertion that research is relatively human-capital- and knowledge-intensive is translated into an extreme specification in which only knowledge and human capital are used to produce new designs or knowledge. Labor and capital do not enter at all. These kinds of restrictions will reduce the analysis of the dynamics of this system to a system of equations that can be explicitly solved by doing algebra. Presumably, a relaxation of these assumptions that preserves the factor intensity orderings used here would not change the basic dynamics of the model.

The set of possible producer durables is indexed by the non-negative half line \mathbb{R}_+ . An input list is therefore a non-negative function x from \mathbb{R}_+ into \mathbb{R}_+ with the property that the integral $\int_0^\infty x(i)di$ is well defined and finite. Let \mathcal{X} denote the set of all such functions. (Formally, \mathcal{X} is the positive cone in the space of absolutely integrable Lebesgue measurable functions.) Final goods output is defined to be a function $Y: \mathbb{R}_+ \times \mathbb{R}_+ \times \mathcal{X} \rightarrow \mathbb{R}_+$ that maps human capital H_1 , labor L , and a list of durables $x(\cdot)$. (H_1 denotes the stock of human capital used in this production activity; H_2 will denote the stock of human capital devoted to the research sector.) The simplest interesting

functional form for Y is an extension of the Cobb-Douglas production function,

$$(1) \quad Y(H_1, L, x) = H_1^{\alpha} L^{\beta} \int_0^{\infty} x(i)^{1-\alpha-\beta} di.$$

This function is defined on an unusually large space of potential inputs, but it is a well behaved, constant-returns-to-scale production function.

The interpretation of the input list $x(i)$ is that it describes the usage of all of the possible types of capital equipment that could enter into production. Thus, $x(10)$ could describe the number of dump trucks in use, $x(17)$ the number of computers, ... The analysis of the model would be almost exactly the same if the goods $x(\cdot)$ were indexed by the integers rather than by the continuum (and the integral replaced by a sum), but this would force explicit consideration of integer constraints that are avoided in the continuum case. At any time t , only a subset of all possible producer durables will be available for sale. Since $A(t)$ will be used to denote the range of goods for which designs exist at time t , any feasible input list at time t will have $x(i) = 0$ for all i greater than $A(t)$.

Because of the constant returns to scale, output in this sector can be described in terms of the actions of a single, aggregate, price-taking firm. Because producing consumption goods is equivalent to producing capital goods, the evolution of capital is given by

$$(2) \quad \dot{K} = Y - C.$$

For a firm in the intermediate input sector, the production of a new producer durable requires the acquisition of a design from the research sector. This represents a fixed cost in the sense that it must be incurred if any positive amount of a new good is to

be produced, but the cost is otherwise independent of the level of output. With one design, a firm is free to produce an arbitrary number of units of output. Production of these units of output also has a variable cost component; more capital is needed if more units are produced.

To make the functional forms as simple as possible, suppose that x units of a producer durable can be produced with an initial investment in one design and with ηx units of capital. (Strictly speaking, because of the assumption that there is a continuum of goods, these quantities must be measured per unit length of the half line \mathbb{R}_+ . Thus, producing durables on an interval of length ℓ at the level x requires an initial purchase of ℓ units of A and the use of $\ell \eta x$ units of physical capital.) This kind of formulation is close to that used by Judd (1985) in his discussion of patents. He also uses a continuum of goods to study a process whereby new goods are introduced as new patents are produced. The model here corresponds to Judd's discussion of the case where patents are infinitely lived. The main difference is that Judd treats the differentiated goods as consumption goods rather than producer durables and uses a form of exogenous technological change to generate growth.

There are many equivalent institutional arrangements that can support any given equilibrium. As was noted above, research and production of new goods could take place within the same firm, but it is easier to describe the equilibrium if the research department is treated as a separate firm and designs are transferred within the firm for an explicit transfer price. Similarly, the analysis that follows assumes that the firm that buys a design and manufactures a differentiated producer durable rents these durables instead of selling them outright. Nothing would change if the durables were sold or if the producer of the design charged a licensing fee per unit sold instead of selling the design. The analysis is further simplified by assuming that the durables do not depreciate. Adding depreciation would merely add a familiar term to the user cost of capital.

Thus H and L are fixed, and K grows by the amount of forgone consumption. It

remains to specify the process for the accumulation of new designs; that is, for the growth of $A(t)$. As noted above, research output depends on the amount of human capital devoted to research. It also depends on the stock of knowledge that is available to an individual doing research. Thus, if individual i devotes human capital H^i to research and has access to an amount A^i of the total stock of knowledge, the rate of production of new knowledge will be $\delta H^i A^i$ where δ is a productivity parameter.

Although other assumptions about secrecy and property rights could be made, the equilibrium here will be based on the idea that anyone engaged in research has free access to the entire stock of knowledge. This is feasible because the knowledge is a nonrival input. All researchers can take advantage of A at the same time. The output of individual i is therefore $\delta H^i A$. Summing across all individuals engaged in research, the aggregate stock of knowledge evolves according to

$$(3) \quad \dot{A} = \delta H_2 A.$$

This equation captures two substantive assumptions and two functional form assumptions. The first substantive assumption is that devoting more human capital to research leads to a higher rate of production of new designs. The second is that the higher is the total stock of designs and knowledge, the higher is the productivity of an engineer working in the research sector. According to this specification, a college educated engineer working today and one working 100 years ago have the same human capital. The engineer working today is more productive because he or she can take advantage of a larger stock of knowledge that accumulated as large numbers of design problems were solved over the last 100 years.

The two functional form assumptions are that output of designs is linear in each of H_2 and A when the other is held constant. These assumptions, like the exclusion of the

inputs L and $x(\cdot)$, is made for analytical convenience. They could be weakened, at the expense of complicating the dynamic analysis. Any extension, however, should preserve the nonconvexity in equation (3). As emphasized in section 2, replication implies homogeneity of degree one in the rival inputs in any production activity. If there are productive nonrival inputs as well, nonconvexities must be present.

A crucial feature of the specification used here is that knowledge enters into production in two distinct ways. A new design enables the production of a new good that can be used to produce output. A new design also increases the total stock of knowledge, and thereby increases the productivity of human capital in the research sector. By assumption, the benefits of from the first productive role for a design are completely excludable, and the benefits from the second completely nonexcludable. In an overall sense, this makes new designs nonrival inputs that are partially excludable. The equilibrium concept used in the analysis that follows is based on the idea that the owner of a design has property rights over its use in the production of a new producer durable but not over its use in the research sector. If an inventor has a patented design on widgets, no one can make or sell widgets without signing a contract with the inventor. On the other hand, other inventors are free to spend time studying the patent application for the widget. In so doing, they may learn knowledge that helps in the design of widgets. The inventor of the widget has no ability to stop the inventors of widgets from learning from the design of a widget.

In theory and in practice, there is always some ambiguity about what constitutes a design for a new and different good and what constitutes a copy of an existing design. In the model, this ambiguity is artificially resolved by the form of the production function Y . This functional form implies that new goods are never close substitutes for existing goods since all of the producer durables enter into production in an additively separable fashion. (See Pascoa, [1987] for a discussion of this general property of "no neighboring goods.") Although it greatly simplifies the analysis, this is not a realistic feature of the model. In

particular, it rules out the possibility of obsolescence. Nonetheless, the general results here should be robust to more careful modeling of the nature of the interaction between different specialized producer durables inputs. What matters for the results is that the knowledge is a nonrival good that is partially excludable and privately provided.

At the aggregate level, H_1 and H_2 are related by the constraint

$$H_1 + H_2 = H.$$

The interpretation of these equations is that any person can devote human capital to either the final output sector or the research sector. Implicitly this formulation neglects the fact that L and H are supplied jointly. To take the equations used here literally, one must imagine that there are some skilled persons who specialize in human capital accumulation and supply no labor.

To fix notation for prices, let spot prices at any point in time be measured in units of current output and let r denote the interest rate on loans denominated in goods. Let P_A denote the price of new designs, and let w_H denote the rental rate per unit of human capital. Because goods can be converted into capital one-for-one, the spot price for capital is 1 and its rate of return is r . Because of the assumption that anyone engaged in research can freely take advantage of the entire existing stock of designs in doing research to produce new designs, it follows from equation (3) that P_A and w_H are related by

$$w_H = P_A \delta A.$$

Once a design has been produced, a large number of potential suppliers of the new good bid for the right to do so. Each of these firms takes the price P_A for designs, the price of 1 for capital goods, and the interest rate as given, but if it begins production it sets prices to maximize profits. Formally, it helps to let the rental price $p(i)$ for the i 'th

durable lie in the range $\mathbb{R}_+ \cup \{\infty\}$ for its durable good. If no firm produces good i , its price can be understood to be $p(i) = \infty$.

Faced with a price list $\{p(i); i \in \mathbb{R}_+\}$ for all of the producer durables, including infinite prices for the durables that have not been invented yet, the representative final output firm chooses a profit maximizing quantity $x(i)$ for each durable. Because it is a constant-returns-to-scale firm, its input demands are defined only after the scale of operation is pinned down. Let L and H_1 be the total amounts of labor and human capital that are used in the production of final output goods. (The split of total H between H_1 and H_2 remains to be determined.) Given values for H_1 and L , it is possible to derive the aggregate demand for the durables from a maximization problem that is conditional on them:

$$\max_{x \in \mathcal{X}} \int_0^\infty [H_1^\alpha L^\beta x(i)^{1-\alpha-\beta} - p(i)x(i)] di.$$

By differentiating under the integral sign, this leads to an inverse demand function

$$(4) \quad p(i) = (1-\alpha-\beta)H_1^\alpha L^\beta x(i)^{-\alpha-\beta}.$$

(There is an important technical issue about what it means to the final goods producer if prices change on a set of values of i that has measure zero. Because of the symmetry in this model, equation (4) can readily be derived by a limiting argument. For a general discussion of this issue, see Pascoa [1987,1990].)

The demand curve in equation (4) is what the producer of each specialized durable takes as given in choosing the profit-maximizing price to set. Faced with given values H_1 , L , and r , an firm that has already incurred the fixed-cost investment in a design will choose a level of output x to maximize its revenue minus variable cost at every date:

$$\begin{aligned}
(5) \quad \pi &= \max_x p(x)x - r\eta x \\
&= \max_x (1-\alpha-\beta)H_1 \alpha_L \beta x^{1-\alpha-\beta} - r\eta x.
\end{aligned}$$

By assumption, the firm can solve this problem at every point in time because it is free to convert units of durables back into general capital and avoid the interest cost if it decides to rent fewer units. At any point in time, the only fixed cost is the initial expenditure on the design. Consequently, the conventional durable goods monopoly problem does not arise here. In equilibrium, demand for the durables will actually be stationary over time, so no disinvestment will actually take place.

Each producer of specialized durables must rent its output to a large number of final goods producers that can operate at any scale. By assumption, it is not possible for the producer to monitor the use of its durables. As a result, price discrimination is not feasible. The best the firm can do is charge the simple monopoly price.

The decision to produce a new specialized input depends on a comparison of the discounted stream of net revenue and the cost P_A of the initial investment in a design. Because the market for designs is competitive, the price for designs will be bid up until it is equal to the present value of the net revenue that a monopolist can extract. At every date t , it must therefore be true that

$$(6) \quad \int_t^\infty e^{-\int_t^\tau r(s)ds} \pi(\tau) d\tau = P_A(t).$$

If P_A is constant (as it will be in the equilibrium described below) this condition can be put in a more intuitive form. Differentiating with respect to time t yields

$$\pi(t) - r(t) \int_t^\infty e^{-\int_t^\tau r(s)ds} \pi(\tau) d\tau = 0.$$

Substituting in the expression for P_A from equation (6) yields

$$(6') \quad \pi(t) = r(t)P_A.$$

This equation says that at every point in time, the instantaneous excess of revenue over marginal cost must be just sufficient to cover the interest cost on the initial investment in a design. (This formulation of the intertemporal zero profit constraint is taken from Grossman and Helpman [1988].)

To complete the description of the model, it remains to specify preferences and endowments. Consumers have standard, discounted, constant elasticity preferences in continuous time:

$$\int_0^{\infty} U(C)e^{-\rho t} dt, \text{ with } U(C) = \frac{C^{1-\sigma}-1}{1-\sigma} \text{ for } \sigma \in [0, \infty).$$

Consumers are endowed with fixed quantities of labor L and human capital H that are supplied inelastically. At time 0, consumers own the existing durable goods producing firms, and the net revenues of these firms are paid to consumers as dividends. Final goods firms earn zero profits and own no assets, so they can be ignored in the specification of endowments.

An equilibrium for this model will be paths for prices and quantities such that:

- i) consumers make savings and consumption decisions taking interest rates as given;
- ii) holders of human capital decide whether to work in the research sector or the manufacturing sector taking the stock of total knowledge A , the price of designs P_A , and the wage rate in the manufacturing sector w_A as given;
- iii) final goods producers choose labor, human capital, and a list of differentiated durables taking prices as given;

- iv) each firm that owns a design and manufactures a producer durable maximizes profit taking the interest rate and the downward sloping demand curve it faces as given, and setting prices to maximize profits;
- v) firms contemplating entry into the business of producing a durable take prices for designs as given;
- vi) the supply of each good is equal to the demand.

4. Discussion of the Model

A reasonable intuition for the behavior of this model can be inferred by considering the Solow model (1956) in which the evolution of A is given exogenously, and the Uzawa model (1965), in which the evolution of A is determined by the allocation of resources between a research sector and a final goods sector. For a fixed amount of A , and therefore a fixed set of producer durables, the model is almost identical to the Solow model. Because of the symmetry in the model, all the durable goods that are available are supplied at the same level, henceforth denoted as \bar{x} . If they were not, it would be possible to increase profits in the producer durable sector by reducing the output of high output firms and diverting the capital released in this way to low output goods. Since A determines the range of durables that can be produced and since η units of capital are required per unit of durable goods, it is possible to solve for \bar{x} from the equation $K = A\bar{x}$. Then output Y can be written as

$$\begin{aligned}
 (7) \quad Y(H_1, L, x) &= H_1^{\alpha} L^{\beta} \int_0^{\infty} x(i)^{1-\alpha-\beta} di \\
 &= H_1^{\alpha} L^{\beta} A \bar{x}^{1-\alpha-\beta} \\
 &= H_1^{\alpha} L^{\beta} A \left[\frac{K}{\eta A} \right]^{1-\alpha-\beta}
 \end{aligned}$$

$$= (H_1 A)(LA)^\beta (K)^{1-\alpha-\beta} \eta^{\alpha+\beta-1}.$$

The last line of this equation shows that the model behaves just like the neoclassical model with labor and human capital augmenting technological change. In particular, it exhibits the usual diminishing returns to capital accumulation. Given the assumed form of preferences, a fixed level of A will lead to an equilibrium with a steady state in which the level of K is determined by the requirement that the marginal product of capital is equal to the discount rate. If A grew at an exogenously specified exponential rate, the economy would converge to a path where K grows at the same exponential rate as A , just as it does in the Solow model.

The nonconvexity evident in the expression for final output as a function of the primary inputs of the model (H , L , K , and A) is supported in a decentralized equilibrium that relies on monopolistic competition. In contrast, the nonconvexity present in equation (3) describing output of designs is supported through competition with external effects that arise from knowledge spillovers. In each case, the nonconvexity arises because the nonrival good A is an input in production. In final output sector, A matters indirectly because of its effects on the availability of the new $x(\cdot)$ goods. In the research sector, A enters directly.

Both spillovers and price-setting seem essential to capturing the features of knowledge in a model of growth. There is little doubt that much of the value to society of any given innovation or discovery is not captured by the inventor, and any model that missed these spillovers would miss important elements of the growth process. Yet it is still the case that private, profit-maximizing agents make investments in the creation of new knowledge and that they earn a return on these investments by charging a price for the resulting goods that is greater than the marginal cost of producing the goods.

5. Solution of the Model for a Balanced Growth Equilibrium

The strategy for characterizing the model that is followed here is to solve for an equilibrium in which the variables A , K , and Y grow at constant exponential rates. Following the accepted terminology, this will be referred to as a balanced growth equilibrium. The intuition from the Solow model suggests that such an equilibrium will exist if A grows at a constant exponential rate. The intuition from the Uzawa model suggests that it is possible for A to grow at an exponential rate because equation (3) for \dot{A} is linear in A . It will grow at a constant rate if the amount of human capital H_2 that is devoted to research stays constant. Verifying that a balanced growth equilibrium exists therefore reduces to the problem of showing that prices and wages are such that H_1 and H_2 remain constant as Y , K , C , and A grow.

By focusing only on balanced growth paths, the analysis neglects the transient dynamics that arise when the economy starts from a ratio of K to A that differs from the ratio that is maintained along the balanced growth path. It should be a feasible to consider the process of convergence to the balanced growth ratio of K to A using the tools used for studying the Solow model and Uzawa model, but this analysis is not attempted here.

Given results from the previous models, it is easy to guess what the behavior for some of the prices and other quantity variables will be. Since output and investment grow at a constant rate, consumption grows at the same rate. Given the assumption of constant elasticity preferences, this will imply a constant interest rate. Results from my earlier model of differentiated inputs (Romer, 1987) suggest that along the balanced growth path, all the growth in K will be devoted to producing new goods rather than to increasing the quantities of the goods that are already being produced. Thus the balanced growth solution should also have a constant level \bar{x} for all inputs that are produced. Because of

the accumulation of K and A , the wage paid for human capital in the final output sector will grow, but the productivity of human capital in research also grows with A . If the productivity of human capital grows at the same rate in both sectors, H_2 will remain constant if the price P_A for new designs is constant.

Figure 1 illustrates the behavior of the inputs in the model. All of the producer durables that have been designed up to time t are used at the level \bar{x} that indicates the height of the rectangle. The width is the measure of the number of designs or durables in use, $A(t)$. The area $A(t)\bar{x}$ is equal to total capital divided by η . Over time, \bar{x} remains constant, and A grows at a constant exponential rate.

It remains to check that this description of a balanced growth path is consistent with all the equilibrium conditions. The condition determining the allocation of human capital between the final output and research sectors says that the wage paid to human capital in each sector must be the same. In the final output sector, the wage for human capital is its marginal product. In the research sector, the wage is $P_A \delta A$, and human capital receives all the income from this sector. To equalize returns to human capital in both sectors, H_2 must be chosen so that

$$(8) \quad w_H = P_A \delta A = \alpha(H-H_2)^{\alpha-1} L^\beta \int_0^\infty x(i)^{1-\alpha-\beta} di.$$

Implicitly, the allocation of H between the two sectors is constrained by the requirement that H_1 and H_2 be non-negative. If either of these constraints is binding, equation (8) will hold as an inequality.

Substituting the constraint relating H_1 and H_2 into equation (7) and retaining the variable \bar{x} gives

$$(9) \quad Y(H_1, L, \bar{x}) = (H-H_2)^\alpha L^\beta \int_0^\infty x(i)^{1-\alpha-\beta} di$$

$$= (H-H_2)^{\alpha} L^{\beta} A(\bar{x})^{1-\alpha-\beta}.$$

When equations (8) and (9) are combined, the price for new knowledge can then be written as

$$(10) \quad P_A = \frac{\alpha}{\delta} (H-H_2)^{\alpha-1} L^{\beta} (\bar{x})^{1-\alpha-\beta}.$$

For a given value of H_2 , the implied exponential growth rate for A is δH_2 . By equation (9), Y must grow at the same rate as A if L , H_1 , and \bar{x} are fixed. If \bar{x} is fixed, then K must grow at the same rate as A , because total usage of capital is $A\bar{x}\eta$. Let g denote the growth rate of A , Y and K . Since K/Y is a constant, the ratio $\frac{\dot{C}}{\dot{Y}} = 1 - \frac{\dot{K}}{\dot{Y}} = 1 - \frac{\dot{K}}{K} \frac{K}{Y}$ must also be constant. The common growth rate g for all these variables is therefore

$$g = \frac{\dot{C}}{C} = \frac{\dot{Y}}{Y} = \frac{\dot{K}}{K} = \frac{\dot{A}}{A} = \delta H_2.$$

From the form of preferences, it then follows that the constant interest rate for this equilibrium is

$$(11) \quad r = \rho + \sigma \frac{\dot{C}}{C} = \rho + \sigma \delta H_2.$$

Equations (10) and (11) form two of the four equations needed to solve for the four values H_2 , P_A , \bar{x} , and r that should be constant in a balanced growth equilibrium. One of the remaining two equations comes from the zero profit condition for the firms producing durable goods. Rewriting the intertemporal zero profit condition in equation (6'), we have

$$(12) \quad \pi = rP_A.$$

The other equation comes from the first order condition for the profit maximization problem for an individual firm in that industry. From equation (5), the instantaneous profit maximization problem of a firm that produces durables is

$$\pi(H-H_2, L, r) = \max_x (1-\alpha-\beta)(H-H_2)^\alpha L^\beta x^{1-\alpha-\beta} - r\eta x.$$

The first order condition for this problem is

$$(13) \quad r = \frac{1}{\eta}(1-\alpha-\beta)^2(H-H_2)^\alpha L^\beta (\bar{x})^{-\alpha-\beta}.$$

Equations (10), (11), (12), and (13) form the system of equations that can be solved for the four variables H_2 , P_A , \bar{x} , and r . Because of the exclusion restrictions built into the model, this system of equations can be explicitly solved. The details are given in the appendix.

The main result from the solution is the expression for the growth rate as a function of the underlying parameters:

$$(14) \quad g = \frac{\delta H - \Lambda \rho}{\Lambda \sigma + 1},$$

where Λ is the constant

$$(15) \quad \Lambda = \frac{\alpha}{(1-\alpha-\beta)(\alpha+\beta)}.$$

The expression for the growth rate suggests a minor technical restriction. For the integral

in the consumer's preferences to be finite, the rate of growth of current utility $(1-\sigma)g$ must be less than the discount rate ρ . Thus, for $\sigma \in [0,1)$, $\frac{(1-\sigma)\delta H}{\Lambda+1}$ must be less than the discount rate ρ . If this does not hold, the integral can be infinite and some kind of overtaking criterion must be used to describe the behavior of the consumer.

As one would expect from the original Uzawa model, the preference parameters σ and ρ both influence the growth rate. A reduction in σ or in ρ , either of which makes people more willing to substitute consumption today for consumption tomorrow, will lead to faster growth. What may be more surprising is that the growth rate does not depend on the size of the labor force L . This result stands in contrast to the implausible result from my earlier paper (Romer 1987) that growth rates are monotonically increasing in L , and it confirms the conjecture that adding human capital as a separate variable and specifying a more reasonable research technology removes the dependence of the growth rate on L .

The growth rate also is independent of the parameter η that determines the cost in forgone output of producing a unit of a durable. The fact that changes in η do not affect the long run rate of growth of knowledge has implications for policies designed to encourage physical capital accumulation. Like Arrow's model of learning-by-doing (1962a) and my earlier model (Romer, 1986), both of which tie the rate of growth of A to that of K , this model has the property that \dot{A}/A is equal to \dot{K}/K along an equilibrium path. In the earlier models, the ratio of A to K was fixed, so any intervention that increased the accumulation of K also increased the accumulation of A and therefore increased the rate of growth. Here, the ratio of K to A is determined endogenously. If the government offers a subsidy to the accumulation of capital that is financed by a lump sum tax, this has the same effect on growth rates as a reduction in η . (To see a demonstration of this kind of result, see Romer 1989). In the model constructed here, a subsidy will have the effect of increasing the ratio of K to A , but it will not affect the rate of accumulation of A . It therefore will not affect growth rates in the long run.

Another interesting feature of this expression for the growth rate is that it suggests the possibility of stagnation if there is not enough human capital. Strictly speaking, the negative growth rates suggested in the formula will not be observed, for they correspond to choices of H_2 that are negative. Since the growth rate g is equal to δH_2 , equation (14) can be restated as an expression for H_2 in terms of H :

$$(16) \quad H_2 = \frac{H - \rho \frac{\Lambda}{\delta}}{\Lambda \sigma + 1}.$$

If time units are chosen so that δ is equal to 1, g is equal to H_2 , and the plot for each is given in figure 2. If H is too low, the non-negativity constraint on H_2 is binding and growth does not take place. In this case, all of the feasible growth rates for A are too small relative to the discount rate to justify the sacrifice in current output necessary for growth to take place. If H is like income, research in this model is like a luxury good: H_2 increases more than proportionally with increases in H .

This result offers one possible way to explain the wide variation in growth rates observed among countries and for the fact that in some countries growth in per capita income has been close to zero. This explanation is reminiscent of the explanation for the absence of growth in prehistoric time that is offered by some historians and anthropologists: civilization, and hence growth, could not begin until human capital could be freed from the daily struggle for survival. This model cannot offer a complete explanation for these observations because it treats the stock of H (and of L) as given, but it does suggest directions for further work.

The expression from the appendix for P_A is

$$(17) \quad P_A^{\alpha+\beta} = \frac{\Gamma}{\delta \alpha} \left[\frac{\delta \sigma \Lambda H + \Lambda \rho}{\sigma \Lambda + 1} \right]^{\alpha-1} L^\beta \eta^{\alpha+\beta-1}$$

where Γ is the constant

$$\Gamma = [\alpha(1-\alpha-\beta)]^{1-\alpha-\beta}(\alpha+\beta)^{\alpha+\beta}.$$

From this it follows that although η and L do not affect the growth rate, they do affect the relative price of designs; an increase in L or a decrease in η causes an increase in P_A . This causes the producer durable sector to substitute away from A towards K , as indicated by an increase in \bar{x} . The expression for \bar{x} in terms of P_A is

$$(18) \quad \bar{x} = \frac{1}{\eta} P_A \frac{1-\alpha-\beta}{\alpha+\beta}.$$

An increase in L or a decrease in η will therefore cause both \bar{x} and the ratio of K to A to increase.

The effects of a subsidy to capital can be contrasted with a policy designed to encourage research. A subsidy to employment in the research sector that is financed through lump-sum taxes has the same effects on growth as an increase in the productivity parameter δ in equation (3). In the long run, this will cause an increase in the growth rate, a fall in P_A , and a reduction in \bar{x} and in the ratio of K to A .

Because of the externality associated with the production of knowledge, there is a presumption that this kind of subsidy to the accumulation of A will improve welfare, but demonstrating this result rigorously forces the analysis to depart from consideration of balanced growth paths. Any intervention designed to move an economy from one balanced growth path to another must consider the transition dynamics along the way, and an explicit analysis of these dynamics is beyond the scope of this paper.

It is easy, however, to compare growth rates along the balanced growth path that emerges from the equilibrium and the one that would emerge from the solution to a social planning problem. The symmetry arguments that lead up to equation (7) imply that the

social planning problem for this economy can be written as:

$$\begin{aligned} \max \int_0^{\infty} \frac{C^{1-\sigma}-1}{1-\sigma} e^{-\rho t} dt \\ \text{s.t. } \dot{K} &= \eta^{\alpha+\beta-1} A^{\alpha+\beta} H_1^{\alpha} L^{\beta} K^{1-\alpha-\beta} - C \\ \dot{A} &= \delta H_2 A \\ H_1 + H_2 &\leq H. \end{aligned}$$

As is shown in the appendix, the balanced growth solution to the first order necessary conditions for this problem has a growth rate g^* that is given by

$$(19) \quad g^* = \frac{\delta H - \Theta \rho}{\Theta \sigma + (1 - \Theta)},$$

where $\Theta = \frac{\alpha}{\alpha+\beta}$. Since Θ is less than the corresponding coefficient Λ in equation (14) and since $1-\Theta$ is less than 1, it follows that the socially optimal growth rate is greater than the growth rate in the decentralized equilibrium. In the social optimum, more human capital is devoted to research and less to the production of final output goods.

6. Growth, Trade, and Research

The final observations about this model pertain to its implications for growth, trade, and research. These can be seen most simply by comparing the balanced growth equilibrium for two identical closed economies that operate in isolation with the balanced growth equilibrium that would obtain if the economies had always been fully integrated. In isolation, the common growth rate is given by g from equation (16), with H set equal to the amount of human capital in each country. In the second case, the growth rate is

found by replacing H by total worldwide human capital, $2H$. Growth goes up, as does the fraction H_2/H of total worldwide human capital devoted to research.

This thought experiment suggests why a decision to engage in trade may be important even for very large economies like that of China or India. If access to a large number of workers or consumers were all that mattered, having a large population would be a good substitute for trade with other nations. The model here suggests that what is important for growth is integration not into an economy with a large number of people but rather into one with a large amount of human capital. Many of the details of trade between different economies of this kind remain to be worked out,² but based on the observation that growth does seem to be correlated with the degree of integration into worldwide markets but not closely related to population size or density, the results from this model seem promising.

Finally, the most direct test of the implications of the model would come from a controlled experiment in which the level of research activity was monitored both before and after a country was opened to trade with the rest of the world. Sokoloff (1988) reports historical data on a natural experiment that comes close to this test. He finds cross-sectional variation that is supportive of the model; counties in the the United States in the early 19'th century that had access to navigable waterways had higher rates of patenting than counties that did not. More convincingly, he shows that over time, the introduction of a water transportation (either because of the construction of a new canal or the dredging of a river) is followed by a sharp increase in the rate of patenting in counties adjacent to the waterway.

In subsequent work with Zorina Khan (1989), he examines the time series variability in patenting for specific individuals to show that there was a substantial group of people with broad general knowledge who moved in and out of the research in response

²Since the first draft of this paper was written, Grossman and Helpman have explored detailed models of trade and growth with this kind of underlying technology (1989a, 1989b).

to aggregate disturbances (primarily business cycles and a trade embargo by the British) just as one would expect if human capital that could be used to do research had alternative uses in manufacturing and commercial trade. They conclude that at least in the 19th century, there was a fairly elastic short-run supply of human capital for use in research. Combined with the prior evidence that patenting activity does respond to changes in the size of the market, this offers some assurance that the basic mechanisms described in the model are relevant for historical experience.

Appendix

A1. Calculation of the Balanced Growth Equilibrium

Equations (10), (11), (12), and (13) can be solved for the variables \bar{x} , P_A , r , and H_2 , by first substituting out the variables \bar{x} and H_2 , leaving two equations in P_A and r . By inserting equation (13) into the expression (5) for π , equation (12) can be written as

$$(A.1) \quad \bar{x} = P_A \frac{1-\alpha-\beta}{\eta(\alpha+\beta)}.$$

Solving equation (11) for H_2 in terms of r yields,

$$(A.2) \quad H_2 = \frac{r-\rho}{\sigma\delta}.$$

Substituting the expressions from equations (A.1) and (A.2) for \bar{x} and H_2 into equation (13) gives P_A implicitly as a function of r :

$$(A.3) \quad P_A^{\alpha+\beta} = \frac{\Omega}{r} L^\beta \left[H - \frac{r-\rho}{\sigma\delta} \right]^\alpha \eta^{\alpha+\beta-1}.$$

The constant Ω in this expression is a function of the exponents α and β :

$$(A.4) \quad \Omega = (1-\alpha-\beta)^{2-\alpha-\beta} (\alpha+\beta)^{\alpha+\beta}.$$

The second equation for $P_A^{\alpha+\beta}$ in terms of r follows by substituting the expressions for H_2 and \bar{x} into equation (10), the equation describing equality of returns to human capital in both the final output and research sectors. The result can be written as

$$(A.5) \quad P_A^{\alpha+\beta} = \frac{\Gamma}{\delta} L^\beta \left[H - \frac{r-\rho}{\sigma\delta} \right]^{\alpha-1} \eta^{\alpha+\beta-1},$$

where the constant Γ is given by

$$\Gamma = \alpha(1-\alpha-\beta)^{1-\alpha-\beta} (\alpha+\beta)^{\alpha+\beta-1}.$$

Equating equations (A.4) and (A.5) and simplifying gives the expression

$$(A.6) \quad r = \frac{\sigma \delta H + \rho}{\sigma \Lambda + 1},$$

where Λ is as defined in the text. Using the relationship between r and H_2 to solve for H_2 , then multiplying to get $g = \delta H_2$ gives equation (14) in the text,

$$g = \frac{\delta H - \Lambda \rho}{\Lambda \sigma + 1}.$$

Substituting equation (A.6) into equation (A.5) and simplifying gives equation (17) in the text,

$$P_A^{\alpha+\beta} = \frac{\Gamma}{\delta \alpha} \left[\frac{\delta \sigma \Lambda H + \Lambda \rho}{\sigma \Lambda + 1} \right]^{\alpha-1} L^\beta \eta^{\alpha+\beta-1}.$$

Equation (18) in the text follows from equation (A.1).

A2. Calculation of the Balanced Growth Social Optimum

To derive the necessary conditions for the social optimization problem PS, construct the current valued Hamiltonian:

$$\mathcal{H} = \frac{C^{1-\sigma}}{1-\sigma} + \lambda[\eta^{\alpha+\beta-1} A^{\alpha+\beta} (H-H_2)^{\alpha} L^{\beta} K^{1-\alpha-\beta} - C] + \mu \delta H_2 A.$$

The necessary conditions follow by maximizing \mathcal{H} with respect to the control variables C and H_2 , and from the equations for the evolution of the multipliers λ and μ :

$$\dot{\lambda} = \rho\lambda - \frac{\partial \mathcal{H}}{\partial K}, \quad \dot{\mu} = \rho\mu - \frac{\partial \mathcal{H}}{\partial A}.$$

The first order condition for maximizing \mathcal{H} with respect to C gives the usual expression relating marginal utility and the multiplier λ :

$$(A.7) \quad C^{-\sigma} = \lambda.$$

If the symbol Δ is used to represent the term $\eta^{\alpha+\beta-1} A^{\alpha+\beta} (H-H_2)^{\alpha} L^{\beta} K^{1-\alpha-\beta}$ from the Hamiltonian, the first order condition for maximizing \mathcal{H} with respect to H_2 can be written as

$$(A.8) \quad \Delta = (H-H_2) \frac{\delta \mu}{\alpha \lambda} A.$$

Then using equation (A.8), the evolution equation for μ can be simplified to yield

$$(A.9) \quad \frac{\dot{\mu}}{\mu} = \rho - \delta \left[\frac{\alpha+\beta}{\alpha} H - \frac{\beta}{\alpha} H_2 \right].$$

For a balanced growth equilibrium, it must be the case that $\frac{\dot{\mu}}{\mu} = \frac{\dot{\lambda}}{\lambda}$ and that $\frac{\dot{C}}{C} = \frac{\dot{A}}{A}$. Using equation (A.7), these can be combined to yield $\sigma \frac{\dot{A}}{A} = \frac{\dot{\mu}}{\mu}$. Then combining

the evolution equation for A , $\frac{\dot{A}}{A} = \delta H_2$, with equation (A.9) gives an equation in H , H_2 , and the basic parameters of the model:

$$(A.10) \quad \delta H_2 = \rho - \delta \left[\frac{\alpha + \beta}{\alpha} H - \frac{\beta}{\alpha} H_2 \right].$$

Equation (19) in the text can be derived from this equation by solving for H_2 and using the fact that the growth rate g is given by $g = \delta H_2$.

REFERENCES

- Abromowitz, Moses. "Resource and Output Trends in the U.S. since 1870." American Economic Review 46, (March, 1956): 5–23.
- Arrow, Kenneth J. "The Economic Implications of Learning by Doing." Review of Economic Studies 29 (June 1962a): 155–173.
- _____. "Economic Welfare and the Allocation of Resources For Invention." In The Rate and Direction of Inventive Activity. Richard Nelson ed. Princeton: NBER and Princeton University Press, 1962b.
- Becker, Gary S. Murphy, Kevin M., Tamura, Robert. "Economic Growth, Human Capital, and Population Growth." Journal of Political Economy, this issue.
- Bornes, Richard and Sandler, Todd. The Theory of Externalities, Public Goods, and Club Goods. Cambridge: Cambridge University Press, 1986.
- Dasgupta, Partha, and Joseph Stiglitz. "Learning-By-Doing, Market Structure, and Industrial and Trade Policies." Oxford Economic Papers 40, (June 1988): 246–268.
- Dixit, Avinash, and Stiglitz, Joseph. "Monopolistic Competition and Optimum Product Diversity." American Economic Review 76 (June 1977): 389–405.
- Thier, Wilfred J. "National and International Returns to Scale in the Modern Theory of International Trade." American Economic Review 72, (June 1982): 389–405.

Gollop, Frank M. and Dale W. Jorgenson. U.S. Productivity Growth by Industry. In New Developments in Productivity Measurement and Analysis, John W. Kendrick and Beatrice N. Vaccara (eds.) Chicago: NBER and University of Chicago Press, 1980.

Griliches, Zvi, "Issues in Assessing the Contribution of Research and Development to Productivity Growth." The Bell Journal of Economics 10 (Spring 1979): 92-116.

Grossman, Gene M. and Helpman, Elhanan. "Product development and international trade." NBER Working paper No. 2809, 1988.

_____. "Comparative Advantage and Long Run Growth" NBER Working paper No. 2540, 1989a.

_____. "Growth and Welfare in a Small Open Economy." NBER Working paper No. 2970, 1989b.

Heckman, James J. "A Life Cycle Model of Earnings, Learning, and Consumption." Journal of Political Economy 84 (August 1976): S11-S44.

Helpman, Elhanan and Krugman, Paul. Market Structure and Foreign Trade. Cambridge: MIT Press, 1985.

Jones, Larry and Manuelli, Rodi. "A Model of Optimal Economic Growth." Stanford University working paper, 1988.

Judd, Kenneth L., "On the Performance of Patents," Econometrica 53 (May 1985):

567–85.

Kendrick, John. "Productivity Trends: Capital and Labor." Review of Economics and Statistics 39 (1956): 248–57.

King, Robert G. and Sergio Rebelo, "Business Cycles with Endogenous Growth," University of Rochester working paper, 1987.

Maddison, Angus. Phases of Capitalist Development, Oxford University Press: Oxford, 1982.

Nordhaus, William. Invention, Growth, and Welfare. Cambridge MA: MIT Press, 1969.

Lucas, Robert E. Jr., "On the Mechanics of Economic Development." Journal of Monetary Economics 22 (1988): 3–42.

Pascoa, Mario. "Monopolistic Competition and No-Neighboring-Goods." CARESS working paper #86-14, University of Pennsylvania, 1986.

_____. "Noncooperative Equilibrium and Chamberlinian Monopolistic Competition." Forthcoming Journal of Economic Theory, 1990.

Rebelo, Sergio. "Long Run Policy Analysis and Long Run Growth." University of Rochester working paper, 1988.

Romer, Paul M. "Increasing Returns and Long-Run Growth." Journal of Political Economy 94 (October 1986): 1002–1037.

_____. "Growth Based on Increasing Returns Due to Specialization." American Economic Review 77 (May 1987): 56-62.

_____. "Capital Accumulation in the Theory of Long Run Growth." Modern Business Cycle Theory, Robert Barro (ed.), Cambridge MA: Harvard University Press, 1989.

Rosen, Sherwin, "A Theory of Life Earnings." Journal of Political Economy 84 (August 1976): S45-S68.

Schumpeter, Joseph. Capitalism, Socialism, and Democracy. New York: Harper Brothers 1942.

Shell, Karl. "Toward a Theory of Inventive Activity and Capital Accumulation." American Economic Review 56 (May 1966):62-8.

_____. "A Model of Inventive Activity and Capital Accumulation," in Essays in the Theory of Optimal Economic Growth, Karl Shell, ed. Cambridge, Mass: MIT Press, 1967.

_____. "Inventive Activity, Industrial Organization, and Economic Growth." In Models of Economic Growth, Mirrlees, J. and Stern, N. eds. New York: John Wiley and Sons, 1973.

Sokoloff, Ken. "Inventive Activity in Early Industrial America: Evidence From Patent Records, 1790-1846." Journal of Economic History 48 (December 1988): 813-850.

_____ and Khan, Zorina. "The Democratization of Invention During Early Industrialization: Evidence From the United States, 1790–1846." University of California at Los Angeles working paper, 1989.

Solow, Robert M. "A Contribution to the Theory of Economic Growth." Quarterly Journal of Economics 70 (February 1956): 65–94.

_____. "Technical Change and the Aggregate Production Function." Review of Economics and Statistics, 39 (1957): 312–320.

Jzawa, Hirofumi. "Optimum Technical Change in an Aggregative Model of Economic Growth." International Economic Review 6 (January 1965): 18–31.

Wilson, Robert. "Informational Economies of Scale." Bell Journal of Economics 6 (Spring 1975): 184–195.

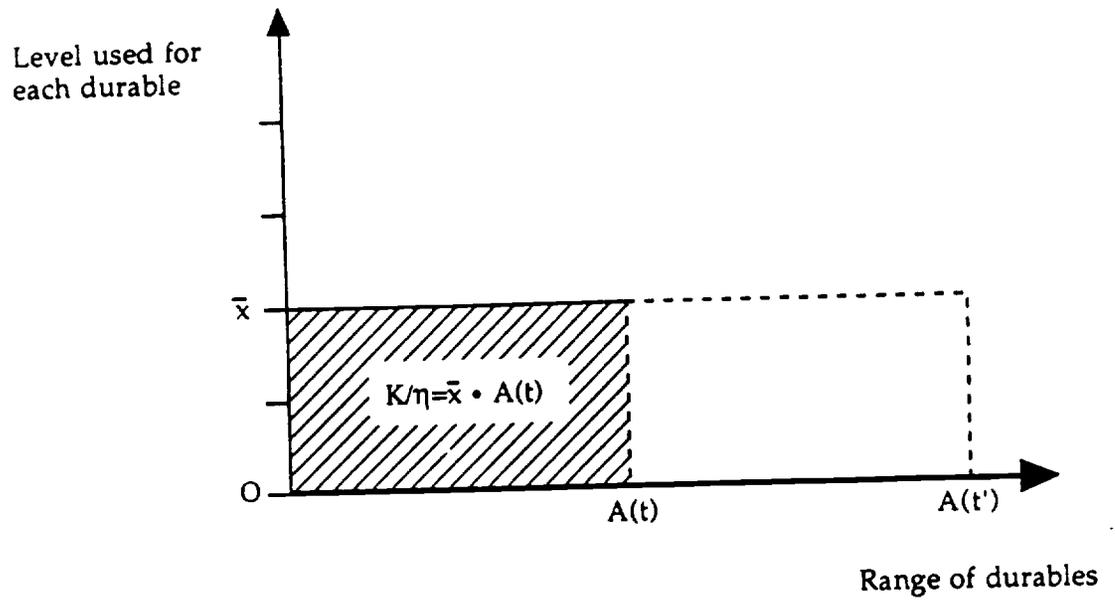


Figure 1: Producer durables used in production at time t and $t' > t$.

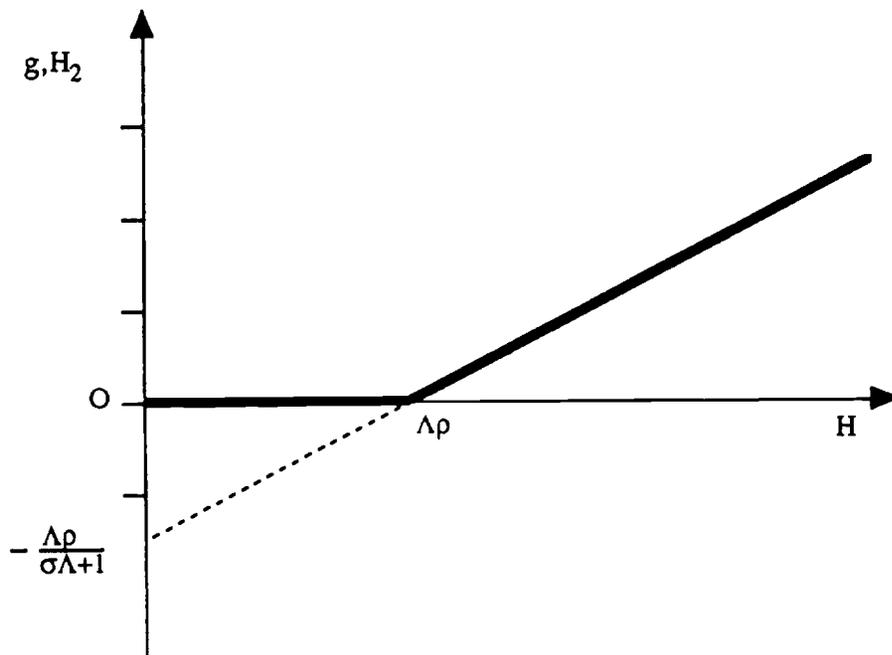


Figure 2: Growth rate and amount of human capital in research as a function of total human capital (for $\delta=1$).