Growth Based on Increasing Returns Due to Specialization

By PAUL M. ROMER*

This note describes an attempt to model increasing returns that arise because of specialization. The idea that increasing returns and specialization are closely related is quite old, but, apparently for technical reasons, we have no fully worked out dynamic model of growth along these lines. There are now several models of growth that consider increasing returns that arise from the accumulation of knowledge. (See, for example, my dissertation, 1983, and 1986a paper; Robert Lucas, 1985; Edward Prescott and John Boyd, 1987.) Despite the presence of aggregate increasing returns, these models can support a decentralized competitive equilibrium with externalities; the externalities arise because of spillovers of knowledge. At least since the publication of Kenneth Arrow's 1962 paper on learning by doing, it has been clear that a competitive equilibrium with externalities provides a tractable framework for the study of increasing returns in a dynamic model. The model described here shows that a closely related framework can be used to study specialization.

The idea that specialization could lead to increasing returns is as old as economics as a discipline. The idea that a decentralized equilibrium with externalities could exist despite the presence of aggregate increasing returns is as old as the notion of an externality. In *Principles of Economics*, Alfred Marshall introduces the notion of an "external economy" to justify the use of a decentralized, price-taking equilibrium in the presence of aggregate increasing returns. He notes in passing that an increase in "trade-knowledge" that cannot be kept secret represents a form of external economy (p. 237). He gives more emphasis to the growth of subsidiary trades that use "machinery of the most highly specialized character" (p. 225), claiming that these too give rise to some vague sort of external effect. In the spirit of specialized endeavors, the model presented below ignores increasing returns from investments in knowledge and external effects due to spillovers of knowledge. It focuses exclusively on the role of specialization. A more realistic and more ambitious model would examine both effects.

*Department of Economics, University of Rochester, Rochester NY, 14627.

I. Static Models of Specialization

The first step in the construction of a model where specialization leads to a form of increasing returns has been taken by Wilfred Ethier (1982). He suggests that we reinterpret as a production function the utility function used by Avinash Dixit and Joseph Stiglitz (1977) to capture a preference for variety. In this reinterpretation, the output of final consumption goods is an increasing function of the total number of specialized intermediate inputs used by a final goods producer. In a continuum version of this model, the list of intermediate inputs used in final good production is a function $x\colon \mathbb{R}_+ \to \mathbb{R}$, where $x(i)$ denotes the amount of intermediate good i used. A production function using both labor and intermediate inputs that is analogous to the Dixit-Stiglitz utility function is

$$(1) \qquad Y(L, x) = L\int_{\mathbb{R}_+} g\left(\frac{x(i)}{L}\right) di,$$

where g is an increasing, strictly concave function with $g(0) = 0$. In the special case considered by Dixit-Stiglitz and by Ethier, g is the power function $g(x) = x^{\alpha}$, with $0 < \alpha$

<1. Then Y takes on the more familiar form

$$(2) \qquad Y(x) = L^{1-\alpha}\int_{\mathbb{R}_+} x(i)^{\alpha}\,di.$$

Let $\{N, M\}$ denote the list of inputs $x(i)$ that takes on the constant value $x(i) = N/M$ on the range $i \in [0, M]$. Thus, M measures the range or number of intermediate inputs used, and N measures the total quantity of such inputs. The graph of $x(i)$ is a rectangle of width M lying on the i axis and having a total area equal to N. In general,

$$(3) \quad Y(L, \{N, M\}) = LMg(N/LM).$$

If g is a power function, this becomes

$$(4) \quad Y(L, \{M, N\}) = M^{1-\alpha}(L^{1-\alpha}N^{\alpha}).$$

In either case, it is easy to show that output of the final good increases with M, the range or number of different inputs, when labor and the total quantity of intermediate inputs are held constant. This loosely captures the idea that a *ceteris paribus* increase in the degree of specialization increases output. In equation (4), Y appears to exhibit increasing returns to scale, but N and M are not the relevant inputs. As a function of labor L and the lists of intermediate inputs $x(i)$, Y is a concave production function that is homogeneous of degree 1.

To capture the idea that fixed costs limit the degree of specialization, assume that the intermediate inputs $x(i)$ are produced from a primary input Z according to a cost function that has a U-shaped average cost curve. Preserving the symmetry in the model, assume that an amount $x(i)$ of any good i can be produced at a cost $h(x(i))$. Inaction at zero cost is feasible, so $h(0)$ equals zero; but at any positive level of production, $h(x)$ is greater than some quasi-fixed cost $\bar{h}$. For simplicity, I assume that this cost is measured purely in terms of the primary input and ignore labor inputs in the production of intermediate inputs. Since this cost is measured in units of the primary good per unit of infinitesimal length di, the resource constraint faced by the economy as a whole is

$$(5) \qquad \int_{\mathbb{R}_+} h(x(i))\,di \le Z.$$

With this specification for costs, the feasible range of intermediate inputs is finite.

Together, a production function like Y and a cost function like h offer an extremely crude representation of the many specialized goods that are in fact used in multiple stages of production. It is intended only as a kind of reduced form. (See Spyros Vassilakis, 1986, for an alternative, more detailed model of specialization.) Modeling the output of a firm in the consumption goods sector as a deterministic function of the entire set of available specialized inputs is a convenient simplification that cannot be taken literally. Besides allowing for multiple stages of intermediate inputs, a more realistic approach would extend this model in precisely the way that Michael Sattinger (1984), Jeffrey Perloff and Steven Salop (1985), and Oliver Hart (1985) extend the Dixit-Stiglitz model of consumer preferences, allowing for many producers of final goods, each of whom has a technology that is most productive with a specific, small subset of all potential intermediate inputs. If the particular inputs that are most productive are distributed symmetrically across a large number of firms producing the final good, the aggregate effect should be similar to that achieved in the model here. If one allows for the possibility of household production, the model can accommodate an apparent preference for variety on the part of consumers as well. (Kenneth Judd, 1985, Nancy Stokey, 1986, and James Schmitz, 1986, are examples of dynamic models with preferences similar to the production function used here.) Ski boots and screw drivers have as much claim to be called intermediate inputs as pig iron and petrochemicals.

A decentralized equilibrium for this economy consists of a continuum of firms in the intermediate goods sector and an indeterminant number of firms producing final output goods with the constant returns to scale production function Y. The final goods firms are assumed to be price takers in all of their markets. Each of the intermediate input producing firms is the single producer of a particular intermediate input and has power in the market for its specialized good. It is still a price taker in the market for the primary input Z. Using final output goods as numeraire, let R denote the price of a unit of the resource Z. (The notation R will more appropriate in the next section where Z is a durable stock in a dynamic model and R has the interpretation of a rental rate.) Assuming for simplicity that the primary input has no alternative use in consumption or production, preferences can be any increasing function of final good consumption. For now, all that I need to specify about the demand side of the economy is that the individual consumers are price takers, and that they are endowed with the stock of the primary resource and an inelastically supplied quantity of labor.

The kind of equilibrium that obtains is a monopolistically competitive equilibrium similar to the one described by Dixit and Stiglitz. Given a list of prices $p(i)$ for the intermediate inputs that are produced, it is straightforward to derive demands for these inputs. Setting the aggregate supply of labor L equal to 1, the (inverse) demand function for any particular input i is proportional to the derivative of the function g that appears as the integrand in Y:

$$p(i) = g'(x(i)). \tag{6}$$

Potential and actual producers of intermediate goods maximize profits taking these demand curves and the price R for the primary resource as given. (My 1986b paper describes a sequence of finite economies that rationalize this as a limit equilibrium.) In equilibrium, some goods i are produced, others are not. All firms in the intermediate goods industry (both potential producers and actual producers) earn zero profits. Given the derived demand curves, profit maximization on the part of intermediate goods producers leads to values of $x(i)$ that depend on the price of the primary resource R. The price R is determined by the requirement that profits for the intermediate goods producers must be zero.

For given Z, the key quantities to be determined are M, the number or range of intermediate inputs that are produced, and $\bar{x}$, the amount of each of these inputs that is produced. By the symmetry in the model, it is clear that all goods that are produced will be produced at the same level. To illustrate the equilibrium in a particular case, let g be the power function described above, and let the cost function h take the form $h(x) = (1 + x^2)/2$. Then the equilibrium quantities are

$$x(i) = \bar{x} = (\alpha/(2-\alpha))^{1/2}, \tag{7}$$

on a set of inputs i of length

$$M = Z(2-\alpha), \tag{8}$$

with $x(i) = 0$ otherwise. The equilibrium value of R can be explicitly calculated, but is not revealing.

It is also straightforward to calculate the quantities that would be chosen by a social planner who maximizes output subject to the constraints imposed by the technology. A curious feature of the choice of g as a power function is that the quantities from the first-best social optimum coincide with those in the decentralized equilibrium. This result relies crucially on the fact that the stock of Z is given. Explicit calculation shows that in the equilibrium, the marginal value of an additional unit of the resource Z is R/α, strictly bigger than the market price, R. In any extension of this model that allows an alternative use for Z, the decentralized equilibrium will differ from the first-best social optimum. In particular, any model that explains growth by allowing individuals to forego current consumption and accumulate additional units of the resource Z will necessarily have an equilibrium with less accumulation of Z than would be socially optimal.

Even with a given quantity of the primary resource Z, a different choice of the function g can lead to equilibrium values for $\bar{x}$ and M that differ from the values that would be chosen by a social planner. The suboptimality arises for two distinct reasons. The downward-sloping demand curve faced by actual producers of intermediate goods causes the equilibrium level of $\bar{x}$ to be too small (and therefore causes M to be too big.) An opposing effect arises because the introduction of a new intermediate input creates surplus for the producers of final goods that cannot be captured by the firm selling the input. New intermediate inputs are introduced up to the point where total costs equal payments to a firm producing an intermediate input, but under standard monopoly pricing these payments are smaller than the surplus created by the additional inputs. This effect causes M to be too small (and therefore causes $\bar{x}$ to be too big.) The case where the function g is a power function happens to be such that these two effects on the quantities $\bar{x}$ and M exactly cancel. However, both effects cause Z to be undervalued.

To highlight the divergence between the private and social gains from the introduction of new goods, it is useful to consider an example that removes the usual distortion arising from a divergence between price and marginal cost. To preserve the result that final output depends nontrivially on the range of inputs used, the function g must have some degree of curvature. Since the derived demand curve for an intermediate input curve is proportional to the derivative of g, this implies that demand must be downward sloping in some region. To insure that price equals marginal cost, the intermediate goods producer must face a demand curve that is horizontal in the relevant region.

Thus, suppose that the function g is at least twice continuously differentiable with the following properties. On the interval $[0, x_0]$, g is strictly concave, with $g(0) = 0$, $g'(x_0) = 1$. On the interval $[x_0, \infty)$, let g have a constant slope equal to 1. In the graph of g, let G denote the intercept that is defined by tracing the constant slope of 1 back to the vertical axis. Thus, for $x > x_0$, $g(x) = G + x$. The curvature in the interval $[0, x_0]$ is needed simply to satisfy the requirement that $g(0) = 0$ without violating continuity. The derived inverse-demand curve $p(i) = g'(x(i))$ is a differentiable curve that may or may not have a finite intercept. It is downward sloping on the interval $[0, x_0]$, and takes on the constant value of 1 on $[x_0, \infty)$.

Consider the output from $Y(L, x)$ with this functional form for g. As before, let $\{N, M\}$ denote the rectangular list of inputs with a range of M different specialized inputs each supplied at the level $x(i) = N/M$. If N/M is greater than x_0 (and by choice of a small enough x_0, this will be true for all relevant lists of inputs), the expression for output as a function of N and M is

$$(9) \qquad Y(L, \{N, M\}) = GLM + N.$$

As before, this is increasing in the range of inputs M when total labor L and the total quantity of intermediate inputs N are held constant. With this function and the previous choice of the cost function $h(x) = (1 + x^2)/2$, it is easy to verify the following equilibrium quantities. (As above, set the total quantity of labor equal to 1.) First, guess that the equilibrium price R for the resource Z is equal to 1. Then the marginal cost of additional units of $x(i)$ measured in units of output goods is $Rh'(x) = x$. The assumption that x_0 is small relative to 1 then implies that marginal cost intersects the marginal revenue schedule at the point $(p, x) = (1, 1)$, which lies in the range where the demand curve is flat; hence, marginal revenue coincides with the demand curve at this point. Since the price R for the primary resource is equal to 1, this is also a point on the average cost curve—in fact, it is the point of minimum average cost—so this corresponds to a potential equilibrium. Given that x_0 is small and provided that the demand price $g'(x)$ does not go to ∞ too rapidly as x goes to zero, the U-shaped average cost curve will lie above the demand curve for all other values of x, tangent only at the point $(1, 1)$. If so, this will be the unique monopolistically competitive equilibrium. In this case, the equilibrium list of inputs $x(i)$ takes on the value 1 for a set of inputs i of measure $M = Z$ and is zero elsewhere.

It is also a simple matter to calculate the solutions to the social planning problem for this economy. For this form of the function g, the decentralized equilibrium leads to a range of output goods that is too small relative to that achieved in the first best social optimum. All firms that produce intermediate goods do so up to the point at which the marginal cost equals the marginal product, so there is no force to offset the tendency for the equilibrium to provide too small a range of inputs. Equilibrium output is $Y = Z(G+1)$, but the price of Z is $R = 1$. For this form of the function g as well as for the previous one, the marginal product of Z is greater than its equilibrium price.

II. A Dynamic Model

One simple way to make the static model into a growth model is to allow for the accumulation of the primary resource Z, which is now interpreted as a durable, general purpose capital good. For simplicity, I treat the supply of labor as being exogenous and neglect both a labor-leisure tradeoff and population growth. The specification of intertemporal preferences is conventional,

$$\int_0^{\infty} U(c(t))e^{-\rho t}\,dt. \tag{10}$$

In the examples that follow, I will assume that the utility function $U(c)$ take the isoelastic form

$$U(c) = (c^{1-\sigma}-1)/(1-\sigma), \qquad \sigma \in (0,\infty). \tag{11}$$

For convenience, let there be a continuum of identical consumers indexed on the interval $[0,1]$, each endowed with an amount $Z(0)$ of the initial stock of general purpose capital. So that I can work interchangeably with per capita and per firm quantities, let there be a continuum of firms in the final goods producing sector, also indexed on $[0,1]$, all producing at the same level. (Because of the constant returns to scale in this sector, this is harmless.) Consumers will rent their capital (i.e. their stock of Z) to intermediate goods-producing firms. These firms use it to produce intermediate inputs $x(i,t)$ according to the technology defined by the cost function h, so that the feasible set of intermediate inputs at every point in time is constrained by equation (5). The intermediate inputs can be interpreted either as a flow of nondurable goods produced by the general purpose capital devoted to the production of inputs of type i, or as a service flow from a durable, specialized capital good of type i, that is created by transforming general purpose capital into specialized capital.

Assuming once again that the aggregate supply of labor is equal to 1, each individual in this economy receives per capita output (equal to per firm output) of $Y(1,x)$. This must be allocated between consumption $c(t)$ and investment in additional capital Z. The simplest investment technology is one that neglects depreciation and permits foregone output to be converted one-for-one into new capital. Thus, assume that

$$\dot{Z} = Y(1,x) - c. \tag{12}$$

Without considering the general problem of how to calculate a dynamic equilibrium with monopolistic competition for this model, it is possible to describe equilibria for the specially chosen functional forms considered here. (For a discussion of general methods for calculating equilibria of this type, see my 1986b paper.) Consider first the case described above where $g(x)$ has a slope of 1 for values of x greater than x_0. From the calculation of the static equilibrium with these functional forms, it is clear that the rental rate R (and now it is a true rental rate) on a unit of Z is equal to one unit of consumption goods per unit time. Since one unit of consumption goods can be converted into one unit of capital Z, the price of capital goods in terms of consumption goods must also equal 1. Thus the instantaneous, continuously compounded rate of return on investments in capital goods is 100 percent per unit time. This preserves the values calculated from the static model and makes sense if the unit used to measure time is roughly a decade. The discount rate ρ must also be scaled up to reflect this choice of time units. However, to ensure that growth will take place, the discount rate is assumed to be less than the return to savings; that is, ρ is assumed to be less than 100 percent.

The value for $\bar{x}$ is 1 and the range of goods $M(t)$ is equal to $Z(t)$. Hence, $N(t) = M(t)\bar{x} = Z(t)$. Since output is given by $Y(L, x) = GLM + N$ and L is assumed to take on the constant value 1, output at time t is $Y(t) = Z(t)(G+1)$. For the specified form of preferences, the instantaneous, continuously compounded interest rate on consumption good loans is $\rho + \sigma(\dot{c}/c)$. For this to be consistent with a rate of return of 100 percent on investments in capital, consumption must grow forever at the exponential rate $(1-\rho)/\sigma$. Because output is linear in Z, this is feasible if Z grows at the same exponential rate and consumption is proportional to Z.

To verify that this is an equilibrium, consider the problem faced by a representative consumer. At time t, the consumer will receive labor income equal to $L(\partial Y/\partial L) = GLM(t)$ and rental income on capital equal to $RZ(t)$. The consumer takes the interest rate $R = 1$ as given and takes the path for labor income over time as exogenously given. The consumer chooses how much to consume and the rate of accumulation $\dot{Z}$. Since the total mass of identical consumers is 1, the aggregate rate of accumulation will also equal $\dot{Z}$. Just as in the static model, the equilibrium condition in the market with monopolistic competition is that the range of inputs produced at time t must satisfy $M(t) = Z(t)$.

Each individual consumer takes the path for $M(t)$ as given because it depends on the aggregate savings decisions of all consumers in the economy. In this sense, $M(t)$ behaves just like a positive externality, like a form of anti-smoke. Using the approach described in my 1986a paper for calculating dynamic equilibrium problems with a path like $M(t)$, which atomistic agents take as given but which is endogenously determined, it is easy to verify that the solution to the consumers problem is indeed to choose $c(t)$ and $Z(t)$ so that they grow at the rate $(1-\rho)/\sigma$. (For example, in the logarithmic case $\sigma = 1$, the equilibrium value of $c(t)$ is $c(t) = (G+\rho)Z(t)$. Substituting this and the expression $Y(t) = Z(t)(G+1)$ into equation (12) shows that c and Z grow at the rate $1-\rho$.)

One can verify directly that this equilibrium is suboptimal. Relative to the maximization problem faced by each consumer, a social planner would not take the path of wages or $M(t)$ as given; instead, the planner would take account of the fact that a higher rate of savings leads not only to higher investment income but also higher labor income. The planner would also produce more output for given Z by setting $\bar{x}$ and M at the (first-best) optimal levels rather than at the equilibrium levels. Both these effects cause the first best optimum to have a higher rate of investment and a higher rate of growth. All individuals in this economy could be made better off by a binding agreement to invest and save more than is privately optimal and to subsidize the production of a wider range of goods.

In my related paper (1986b), I argue that it is not an accident that the analysis of this equilibrium so strongly resembles one with a positive externality. This apparent "external economy" associated with the specialization is closely related to the intuition behind Marshall's use of the term. This model is not one with a true positive externality, but it nonetheless behaves exactly as if one were present.

The analysis of the dynamic equilibrium with the same preferences and cost function h, but with $g(x) = x^{\alpha}$ is quite similar. The only important difference is that the equilibrium value of R, while still constant, differs from the previous value of 1. Consumption and the stock of Z will still grow at a constant rate (though one that is algebraically more complicated to express.) The equilibrium is still suboptimal, growing more slowly than the first best optimum. Even though the static equilibrium is efficient for given a level of Z, the dynamic equilibrium offers individual agents a return from savings that is too small, and Z grows too slowly. The only intervention needed to achieve the optimum in this special case is a subsidy to savings.

In both of these equilibria, the economy will behave as if there is a form of exogenous, labor augmenting technological change. In the second case this is easy to compare with standard Cobb-Douglas descriptions of growth. Equations (7) and (8) imply that both $N(t)$ and $M(t)$ are proportional to $Z(t)$. Using output written in terms of L, M, and N as in equation (4), and impounding all the constants into a new constant A, output at time t can be written as

$$(13) \quad Y(t) = M(t)^{1-\alpha}\left(L^{1-\alpha}N(t)^{\alpha}\right)$$
$$= AZ(t)L^{1-\alpha}.$$

In equilibrium, labor's share in total income is $1-\alpha$ and capital's share is α, despite the fact that the true coefficient on Z is 1. A 1 percent increase in the stock of Z causes a 1 percent increase in income, a fraction α of which is returned as payments to capital. The remaining $1-\alpha$ percent increase shows up as increased wages for labor, so labor receives the surplus arising from the apparent increasing returns. Since the rate of return on capital does not decrease with the level of the capital stock, growth can continue indefinitely. Each individual agent takes the path for $M(t)$ as given, so viewed from the aggregate level, the evolution of this economy will appear to be governed by a Cobb-Douglas technology and exogenous technological change. But any change that leads to an increase in savings—for example a tax subsidy, a decrease in the rate of impatience ρ, or a decrease in the intertemporal substitution parameter σ—will cause growth to speed up; the rate of exogenous technological change will appear to increase.

REFERENCES

Arrow, Kenneth J, "The Economic Implications of Learning by Doing," *Review of Economic Studies*, June 1962, *39*, 155–73.

Dixit, Avinash K., and Stiglitz, Joseph, "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, June 1977, *67*, 297–308.

Ethier, Wilfred J., "National and International Returns to Scale in the Modern Theory of International Trade," *American Economic Review*, June 1982, *72*, 389–405.

Hart, Oliver D., "Monopolistic Competition in the Spirit of Chamberlin: A General Model," *Review of Economic Studies*, October 1985, *52*, 529–46.

Judd, Kenneth L., "On the Performance of Patents," *Econometrica*, May 1985, *53*, 567–85.

Lucas, Robert E., Jr., "On the Mechanics of Economic Development," prepared for the Marshall Lecture, May 1985.

Marshall, Alfred, *Principles of Economics*, London: MacMillan Press, 1977.

Perloff, J. and Salop, S., "Equilibrium with Product Differentiation," *Review of Economics Studies*, January 1985, *52*, 107–20.

Prescott, Edward C. and Boyd, John, "Dynamic Coalitions, Growth, and the Firm," in E. C. Prescott and N. Wallace eds., *Contractual Arrangements for Intertemporal Trade*, Vol. 1, Minnesota Studies in Macroeconomics and Time Series Analysis, Minneapolis: University of Minnesota Press, forthcoming 1987.

Romer, Paul M., "Dynamic Competitive Equilibria with Externalities, Increasing Returns, and Unbounded Growth," unpublished doctoral dissertation, University of Chicago, 1983.

______, (1986a) "Increasing Returns and Long-Run Growth," *Journal of Political Economy*, October 1986, *94*, 1002–38.

______, (1986b) "Increasing Returns, Specialization, and External Economies: Growth as Described by Allyn Young," Working Paper No. 64, Rochester Center for Economic Research, December 1986.

Sattinger, M., "Value of an Additional Firm in Monopolistic Competition," *Review of Economic Studies*, April 1984, *51*, 321–32.

Schmitz, James, "Optimal Growth and Product Innovation," Department of Economics Working Paper, University of Wisconsin, October 1986.

Stokey, Nancy L., "Learning-by-Doing and the Introduction of New Goods," Working Paper No. 699, Center for Mathematical Studies in Economics and Management Science, Northwestern University, September 1986.

Vassilakis, Spyros, "Increasing Returns and Strategic Behavior," unpublished doctoral dissertation, Johns Hopkins University, 1986.