IRREVERSIBLE INVESTMENT WITH UNCERTAINTY AND SCALE ECONOMIES

by

Avinash Dixit
Princeton University

Contents:

Abstract
1. Introduction
2. The Model without Scale Economies
3. The Model with Scale Economies
4. External Economies
5. Concluding Comments
Appendix
References
Figures 1 – 5

Discussion Paper
No.TE/92/240
May 1992

The Suntory Centre
Suntory and Toyota International Centres for Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel.: 020-7955 6698

I thank Giuseppe Bertola, Miles Kimball, John Leahy and participants at seminars at Princeton and New York Universities for comments on the Beta-testing version, and the National Science Foundation and the Guggenheim Foundation for financial support.

Abstract

This paper analyses optimal irreversible investment policy when profits are subject to a multiplicative geometric Brownian motion shock. The marginal product of capital is increasing initially and decreasing thereafter. In the latter range, optimal policy is familiar: capacity is added gradually as the shock rises to a threshold where the expected return on the marginal unit is a required multiple of the cost of capital. The multiple reflects the option value of waiting. The optimal policy in the increasing marginal product range obeys the sample multiple, now applied to the total return on the discrete increase in capital. Implications for economic growth, and suboptimal equilibria under external economies, are examined.

Keywords: Scale economies; uncertainty; optimal irreversible investment policy; capital; economic growth; profits.

1 Introduction

The theory of investment decisions of firms has gradually recognized the importance of irreversibility and uncertainty. When the two occur together, the rate of return that justifies investment exceeds the Marshallian normal return because of an option value of waiting. The burgeoning literature on this "real options" view of investment is surveyed by Pindyck (1991), and a non-technical account can be found in Dixit (1992).

A third important feature of many investment projects is indivisibility, or more generally, increasing returns to scale. The literature has generally treated this on its own, assuming away uncertainty. Skiba (1978) gives a complete treatment of the Ramsey problem when the production function has an initial convex portion followed by a neoclassical concave one. Cases where increasing returns last for ever are studied in the new growth theory, for example Romer (1986).

*I thank Giuseppe Bertola, Miles Kimball, John Leahy and participants in seminars at Princeton and New York Universities for comments on the Beta-testing version, and the National Science Foundation and the Guggenheim Foundation for financial support.

There is also some literature on lumpy investment; see Manne (1961) or Weitzman (1970). But its scope is limited to finding the plan of capacity expansion to meet a growing demand (sure or stochastic) at minimum cost. Thus price responsiveness and profit-maximization are not considered.

In this paper I shall examine irreversible investment when uncertainty and increasing returns occur jointly, in the more general framework that the "real options" approach offers. The results are natural extensions of the diminishing returns case. The optimal policy has jumps in the capital stock to cross the region of increasing returns. The thresholds that trigger such jumps, and the sizes of the jumps, are such that the expected *total* return exceeds the Marshallian normal return by the same factor that captures the option value of waiting.

Scale economies can arise for different reasons; external economies in human capital or research have been stressed in recent research on growth theory. I shall briefly discuss the case where each firm sees diminishing returns to its own expansion but creates external economies at the industry level. Here the equilibrium path of investment will be suboptimal in some new and interesting respects.

2 The Model Without Scale Economies

The setting is very similar to that of Pindyck (1988) and Bertola (1989), except for the new feature, namely scale economies. To set the stage and facilitate the comparisons, I shall begin with a brief recapitulation of their model. The state of the firm at any moment is described by two variables, the installed capacity K and a stochastic shift variable P. These determine its operating profit flow, $\pi = P\,F(K)$. This should be thought of as the reduced form after the instantaneous optimization of variable inputs. The shock variable is then a composite of many kinds of randomness: prices of output and variable inputs, and shifts of technology. For brevity and concreteness, I shall refer to P as the price, except when some other particular interpretation is more appropriate. I assume that P follows a geometric Brownian motion

$$dP/P = \mu\,dt + \sigma\,dw\,. \tag{1}$$

The "production function" $F(K)$ has the standard neoclassical properties: $F(0) = 0$, $F'(K) > 0$, $F''(K) < 0$.

The firm can expand capacity at the cost of c per unit, but this cost is sunk and the investment is irreversible. Capital is infinitely lived. The operating profit flow is always non-negative, so there is no reason to suspend or abandon operation.[1] The capacity expansion or investment policy seeks to maximize the expected present value of profits net of capacity cost. The discount rate is given and equal to ρ.

Let the firm start with $P = P_0$ and $K = K_0$; the initial capacity is historically determined and not necessarily optimal given the initial price. The optimal policy from here on is described by an upward-sloping threshold curve $P = P(K)$, shown in Figure 1. In the region above the curve, it is optimal to increase capacity in a lump to move immediately to the threshold curve. In the region below the curve, inaction is optimal. The firm waits until the stochastic process of P moves it vertically to $P(K)$, and then invests just enough to keep from crossing the threshold.[2]

To obtain the function $P(K)$ in terms of the data of the problem, we must begin by constructing an auxiliary quadratic equation

$$q(\xi) \equiv \tfrac{1}{2}\sigma^2 \xi(\xi - 1) + \mu\xi - \rho = 0. \tag{2}$$

To find the roots of this, note that (i) Since $\rho > 0$, we have $q(0) < 0$. But $q(\xi) > 0$ when ξ is negative and large in magnitude. Therefore there is a negative root. (ii) We need $\rho > \mu$ to ensure convergence of expected profits for any given capacity. Then $q(1)$ is negative. But $q(\xi)$ becomes positive when ξ is large and positive, so the other root exceeds unity. Call the latter root β.

The threshold function then takes the form

$$P(K) \equiv \frac{\beta}{\beta - 1}\,\frac{(\rho - \mu)\,c}{F'(K)} \tag{3}$$

Since $F'(K)$ is decreasing, $P(K)$ is increasing, as shown in Figure 1.

The net worth of a firm starting at (K, P) with $P \leq P(K)$, and following the optimal policy thereafter, is given by

$$V(K, P) = B(K)\,P^{\beta} + \frac{P\,F(K)}{\rho - \mu} \tag{4}$$

where the second term is the expected present value of profits if capacity is never increased, and the first term is the value of the firm's future expansion options.

---

[1] Depreciation, partial reversibility, abandonment etc. can be handled as in McDonald and Siegel (1985), Brennan and Schwartz (1985), Dixit (1989) and Bertola (1989). To focus more clearly on the scale economies aspect, I have left all these complications out of this paper.

[2] In technical terms, this is "instantaneous" or "barrier" control at $P(K)$; see Harrison and Taksar (1983). The model can formally be adapted to their framework by defining a new state variable $P\,F'(K)$.

At the threshold $P(K)$, this function satisfies two conditions: value matching

$$V_K(K, P(K)) \equiv B'(K)\, P(K)^{\beta} + \frac{P(K)\, F'(K)}{\rho - \mu} = c, \tag{5}$$

and smooth pasting

$$V_{KP}(K, P(K)) \equiv B'(K)\, \beta\, P(K)^{\beta-1} + \frac{F'(K)}{\rho - \mu} = 0\,. \tag{6}$$

In fact the function $P(K)$ is found by eliminating $B'(K)$ between these two equations.[3]

We get better intuition for the solution by contrasting it to the case without uncertainty. If P is to grow at rate μ with certainty, then the present value of profits contributed by a marginal unit of capacity is $P\, F'(K)/(\rho - \mu)$. The unit will be installed if this exceeds the cost c. For the case with uncertainty, the threshold $P(K)$ defined in (3) satisfies

$$\frac{P(K)\, F'(K)}{\rho - \mu} = \frac{\beta}{\beta - 1}\, c; \tag{7}$$

which is just a multiple $\beta/(\beta - 1)$ of the threshold under certainty. Thus uncertainty raises the threshold price that makes it optimal to add any given unit of capacity. This is the option value effect: by waiting a little the firm can get an additional observation of the price, and reduce its downside risk. If the currently expected return to the marginal investment only just covers the cost, then it is clearly optimal to wait a little. Only when the currently expected return rises sufficiently high does it become optimal to exercise the option to invest. For extensive discussions, see Pindyck (1991) and Dixit (1992).

3 The Model With Scale Economies

Now turn to the main focus of this paper, namely the case of a firm whose production function has some increasing returns. I assume the simplest and most familiar form. There is an initial range of increasing returns followed by diminishing returns; the marginal and average product curves have inverse U-shapes. The total product curve is shown in Figure 2, and the corresponding average and marginal curves in Figure 3. I assume $F(0) = 0$; there may be an initial range over which $F(K) \equiv 0$ although the

[3] To find $B(K)$, we eliminate $P(K)$ and note that the value of further expansion options must go to zero as K goes to infinity. Then

$$B(K) = \left(\frac{\beta - 1}{c}\right)^{\beta-1} \left(\frac{1}{\beta(\rho - \mu)}\right)^{\beta} \int_K^{\infty} F'(k)\, dk.$$

figure does not show this. The marginal product $F'(K)$ is increasing over the interval $[0, K^*]$ and decreasing thereafter; the average product $F(K)/K$ is increasing over the larger range $[0, K^{**}]$ and decreasing thereafter. Each of the Figures 2 and 3 contains several other features to do with the optimal investment policy; their meaning will be elucidated shortly.

If the firm's initial capacity K exceeds K^*, then any further capacity expansion is subject to diminishing marginal returns. The optimal policy can be found as above, and the formula (3) for the threshold function $P(K)$ remains valid. Now consider the optimal policy when the initial capacity is less than K^*. Suppose we formally define $P(K)$ to the left of K^* by the same formula. Over this range it will be a decreasing function since $F'(K)$ is increasing. Figure 4 shows this U-shaped curve $P(K)$. (Other features in the figure will be explained soon.)

Now it might be tempting to think that $P(K)$ will continue to serve as the threshold on its decreasing portion just as it does on the increasing branch. Starting at $K < K^*$, if P rises to $P(K)$, the firm would invest capacity in a lump to move to the right hand branch, and thereafter proceed gradually as in the previous section. But that cannot be right. For example, if $F'(K)$ goes to 0 as K goes to zero, $P(K)$ will go to ∞. But starting with zero capacity, a sufficiently high finite price will make it profitable to install a lump of capital with positive total product.

The instinct, although wrong, is along the right lines. It turns out that there is a downward-sloping curve, shown as $Q(X)$ in Figure 4, that serves the same purpose. Starting at any X less than K^*, the firm does nothing if the price is below $Q(X)$. If this threshold is reached, it invests in a lump to raise the capacity to K along the right hand branch of the curve $P(K)$. The figure shows an example of this.

I will first explain the construction of this curve $Q(X)$ and the economic intuition for the investment policy based on it. The mathematical arguments for the optimality of this policy will follow in a separate subsection.

3.1 Construction of the Policy

In Figure 2, let $K_0 = X < K^*$ be the firm's starting level of capital. From the point $(X, F(X))$ on the production function, draw a line that becomes a tangent to the production function at a point $(K, F(K))$ to the right of K^*. Under our assumptions this is unique. The slope of the line is

$$F'(K) = \frac{F(K) - F(X)}{K - X}. \tag{8}$$

With $K > K^*$, we can read off $P(K)$ from the right hand branch of that curve, or from the formula (3). Define $Q(X) = P(K)$. As X ranges over $(0, K^*)$, this produces a curve that is shown in Figure 4. When the pair of points, X to the left of K^* and K to the right of K^*, are related by the construction above, the corresponding points on the curves $Q(X)$ and $P(K)$ are to have equal heights.

From Figure 2 we see that as X moves gradually to the left from K^*, the corresponding point K moves to the right from K^*, finally reaching K^{**} when X reaches 0. Thus $Q(X)$ is a downward-sloping curve, starting at $Q(0) = P(K^{**})$ and merging with the left hand branch of the $P(K)$ curve as X approaches K^*.

The construction can be seen even more clearly from Figure 3 using the average and marginal products. Write the condition (8) defining the relation between X and K as

$$F(K) - F(X) = F'(K)\,(K - X). \tag{9}$$

The left hand side is just the area under the marginal product curve between X and K. The right hand side is the rectangle with its base from X to K and its height equal to the height of the marginal product curve at K. The condition therefore amounts to equality of the two areas shown shaded in Figure 3. For each K between K^* and K^{**}, we can construct such an X. As K increases over its range, X decreases from K^* to 0. Let $G(X)$ be the curve traced out by the northwest corner of the rectangle whose area is the right hand side of (9). This is an upward-sloping curve starting at $G(0) = F'(K^{**})$. It lies entirely above the marginal product curve, and merges with the latter at K^*. Further,

$$Q(X) = P(K) = \frac{\beta}{\beta - 1}\,\frac{(\rho - \mu)\,c}{F'(K)} = \frac{\beta}{\beta - 1}\,\frac{(\rho - \mu)\,c}{G(X)},$$

so $Q(X)$ is obtained from $G(X)$ in exactly the same way as the right hand branch of $P(K)$ was obtained from $F'(K)$ in (3).

The investment policy starting from $K_0 = X$ and an initial price $P_0 < Q(X)$ is to wait until the price rises to the threshold $Q(X)$, and then invest in a jump to the point K that corresponds to X in the construction above. We can see the economics of this by substituting (8) into (3). We get

$$\frac{P(K)\,[F(K) - F(X)]}{\rho - \mu} = \frac{\beta}{\beta - 1}\,c\,(K - X). \tag{10}$$

The left hand side is the expected discounted present value of the profit flow contributed by the discrete increment of capital from X to K. On the right hand side, $c(K - X)$

is the cost of installing this increment. The condition requires the expected profit gain to be a multiple $\beta/(\beta - 1)$ of the cost. This is an exact analog of the result (7) for the diminishing returns case. All we need to do is to replace a marginal condition by the corresponding total condition.

3.2 Mathematical Argument

The claim is that the investment policy described above is optimal. I will now investigate this somewhat more formally. Starting with a policy of this form, I will verify that it satisfies all the necessary conditions of optimality. Of course in a non-convex problem there can be multiple solutions to the necessary conditions, and I have no formal proof that the policy I find is globally optimal. But its excellent economic sense constitutes a strong argument in its favor. In particular, the option value interpretation, exactly analogous to the conventional diminishing returns case but now based on a total condition rather than a marginal one, is compelling.

Let the starting capital level be $X < K^*$ and the starting price P. Consider a policy defined by two numbers $Q \geq P$ and $K \geq X$ as follows: Wait until the first instant T at which the price P_T rises to equal Q, and then raise the capacity level to K. The expected payoff from this policy is

$$\mathbf{E}\left\{ F(X) \int_0^T P_t\, e^{-\rho t}\, dt + [V(K,Q) - c(K - X)]\, e^{-\rho T} \right\}$$
$$= F(X)\, \mathbf{E}\left\{ \int_0^T P_t\, e^{-\rho t}\, dt \right\} + [V(K,Q) - c(K-X)]\, \mathbf{E}[e^{-\rho T}]$$

where $V(K,Q)$ denotes the continuation value of the firm when optimal policies are followed starting at the new point (K,Q). This function cannot yet be assumed known, although we will see how to evaluate it.

The expected values that appear in the above expression can be found using standard methods as in Karlin and Taylor (1975, chapter 7) or Harrison (1985, chapter 3); the Appendix sketches the details. The results are

$$\mathbf{E}[e^{-\rho T}] = (P/Q)^\beta, \tag{11}$$
$$\mathbf{E}\left\{ \int_0^T P_t\, e^{-\rho t}\, dt \right\} = [P - P^\beta Q^{1-\beta}]/(\rho - \mu), \tag{12}$$

where $\beta > 1$ is the root of the quadratic (2) defined in the previous section.

Substituting, and choosing Q and K optimally, we have the Bellman equation for the initial value of the firm:

$$V(X,P) = \max_{K,Q} \left\{ \frac{F(X)}{\rho - \mu} \left[P - P^{\beta} Q^{1-\beta}\right] + \left(\frac{P}{Q}\right)^{\beta} \left[V(K,Q) - c(K-X)\right] \right\} \quad (13)$$

Assuming for the moment that the constraints $K \geq X$ and $Q \geq P$ are not binding, consider the first-order conditions for maximizing the right hand side (RHS) of (13):

$$\begin{aligned} \frac{\partial RHS}{\partial K} &= \left(\frac{P}{Q}\right)^{\beta} \{V_K(K,Q) - c\} \\ &= 0. \end{aligned} \quad (14)$$

and

$$\begin{aligned} \frac{\partial RHS}{\partial Q} &= \left(\frac{P}{Q}\right)^{\beta} \left\{ \frac{\beta - 1}{\rho - \mu} F(X) + V_Q(K,Q) - \beta \frac{V(K,Q) - c(K-X)}{Q} \right\} \\ &= 0. \end{aligned} \quad (15)$$

Begin by observing that the expression in the brackets in both first-order conditions are independent of P. In particular, the threshold Q depends only on the current capacity X, not on the current price P. Write it as $Q(X)$. This will of course turn out to be the same $Q(X)$ that was defined in the previous subsection, but that remains to be established.

Assume for the moment that $Q(X)$ is a decreasing function, and its value exceeds $P(K^*)$ for all $X < K^*$, as shown in Figure 4. For a given X, when the price rises to this threshold, the level K to which capacity is increased is given by the first-order condition (14), which requires $V_K(K,Q) = c$. This is just the value matching condition (5) that is satisfied when (K,Q) lies along the curve $Q = P(K)$ to the right of K^*. Let us therefore proceed with the policy of setting K along this curve. There the function $V(K,Q)$ is given by (4).

We can use this information in turn in the first-order condition for Q to simplify it further:

$$\begin{aligned} 0 &= \frac{\beta - 1}{\rho - \mu} F(X) + \left[\beta B(K) Q^{\beta - 1} + \frac{F(K)}{\rho - \mu}\right] - \beta \frac{B(K)Q^{\beta} + F(K)/(\rho - \mu) - c(K-X)}{Q} \\ &= \frac{\beta - 1}{\rho - \mu} \left[F(X) - F(K)\right] + \beta c \frac{K - X}{Q}. \end{aligned}$$

Recognizing that $Q = P(K)$ is given by (3), this simplifies further to

$$F(K) - F(X) - F'(K)(K - X) = 0.$$

This is exactly the condition (8) or (9) obtained when I described the claimed optimal policy in the previous subsection. Thus the $Q(X)$ defined in this subsection is the same as that constructed earlier, and all its properties obtained there are available for use. In particular, the mathematical steps in this subsection proceeded on the assumption that $Q(X)$ was a decreasing function, and greater than $P(K^*)$ for all X in the range $(0, K^*)$. These assumptions have now been verified.

The argument is still incomplete. I began by considering policies in a particular class - wait until price reaches Q and then invest in a lump to increase the capital stock to K - and then optimized with respect to the two numbers Q and K. We should verify that the resulting policy satisfies the first-order conditions appropriate to the larger class of all feasible policies, namely ones that at any time use no more information than is then available. The appropriate conditions are developed in Harrison, Sellke and Taylor (1983).

With the definitions above, when $P = Q(X) = P(K)$, a jump from X to K is claimed to be optimal. With $P = Q$, the Bellman equation (13) becomes

$$V(X, P) = V(K, Q) - c(K - X),$$

which is the total condition (1.11) in Harrison et al. We have already established the first-order condition $V_K(K, Q) = c$ at the terminal point of the jump. Differentiating the Bellman equation using the envelope theorem gives

$$V_X(X, P) = (P/Q)^{\beta} c,$$

so when $P = Q$ we have $V_X(X, Q) = c$, which is the first-order condition at the initial point of the jump. The two first-order conditions together constitute (1.10) in Harrison et al. Thus all the necessary conditions are verified.

3.3 Discussion

Subject to the warnings about multiple local optima, we now have the following optimal policy. In Figure 4, the region below the combined curve formed by $Q(X)$ for $X < K^*$ and $P(K)$ for $K > K^*$ is the region of inaction. Starting at a point here, the firm waits until the stochastic movement of the price takes it to a point on the curve vertically above its initial capacity. Then, from an initial point on the left branch of the combined curve, the firm jumps horizontally to the corresponding point on the right hand branch. From any point above the combined curve, an immediate jump to the right hand branch is similarly made. Once on the right hand branch, further infinitesimal increments to capacity are made as necessary, to prevent crossing above the threshold curve whenever a price rise causes the curve to be hit.

If the firm starts without any installed capacity at all ($X = 0$), it should wait until the price reaches the level $P(K^{**})$, and then jump all the way to K^{**} where the average product is maximum. Thus the firm should never operate in a region where increasing returns are not fully exhausted. But a previous error, or some unexpected change in technology, may leave the firm with a positive initial capacity below K^{**}. If the initial capacity X is below K^{*}, then a jump is justified when the price reaches a threshold $Q(X)$ lower than $P(K^{**})$, and the jump leaves the firm short of K^{**}. Both these observations call for comment. The explanation of the lower threshold is that the firm already has the first few units of capacity with the smallest marginal product. The next ones it installs will have higher marginal products, therefore a smaller price is enough to justify their installation than if the firm had to start from scratch. But with a smaller price threshold, the jump in capacity must stop at an earlier point of the diminishing returns range; the next increment is not justified unless the price rises further. This is the second observation.

An alternative interpretation is to suppose that the firm regards the initial $X < K^{*}$ and the corresponding output flow $F(X)$ as irrelevant bygones. It regards the new investment say $Z = K - X$ as the choice variable, and starts at $Z = 0$. Let $\phi(Z) = F(K) - F(X)$ be the production function. Figure 2 shows that at the terminal point K of the optimal jump from X, the average product $\phi(Z)/Z$ is maximized. Thus the point serves the same role as K^{**} did in the original problem.

Finally, contrast this model with uncertainty to the corresponding model under certainty. Skiba (1978) has a similar production function, and finds that if the initial capital stock is below a critical level, the optimal policy is never to invest. Since he allows exponential depreciation, the initial capital is optimally allowed to wither away. Uncertainty offers a more optimistic scenario. No matter how low the initial capital stock, the shock variable may reach the threshold, and will do so with probability one if the trend is positive, making it possible to jump over the hurdle of low marginal product.

3.4 Implications for Growth

Let us recapitulate the investment process developed above. There is an initial range of increasing returns to capacity, and the profit function is subject to stochastic shocks.

When the shock rises to a threshold level, investment starts with a jump across the region of increasing returns. In response to further favorable shocks, there are further incremental investments. This picture has some implications for the process of industry evolution and economic growth.

Consider an economy with several output sectors, each of which has some increasing returns, and each is subject to shocks from technical progress, world price changes, or variable input price changes. When the shift variable for one previously dormant sector reaches its threshold, its capacity will increase suddenly. Thereafter it will experience further marginal expansion from time to time. But some other dormant sector may hit its jump threshold. From an aggregate perspective, we will see occasional spurts of growth as a dormant sector expands suddenly. Between a pair of successive spurts, there will be tranquil intervals when all the established sectors add capacity gradually.

This is an appealing view of the growth process, with intriguing parallels to the concept of "punctuated equilibrium" in biological evolution. But the remarks here are only suggestive; a complete theory along these lines will require substantial modelling effort. The different sectors are linked vertically (using others' outputs as inputs) or horizontally (bidding for commonly used scarce inputs or the final buyers' dollars). Thus the shocks for any one sector will have endogenous components in the general equilibrium. It will be necessary to allow decline of some sectors to reallocate resources to growing or spurting sectors. A fully articulated model that incorporates all these links remains a distant research vision.

4 External Economies

In the work so far, the scale economies were internal to the firm; its profit incorporated the effect of its expansion on the marginal and average products. But in many situations of interest, scale economies are external to each firm and occur at the industry level. The firm cannot appropriate the benefit it confers on others, and its investment decisions are suboptimal. The static theory of this is well known, but in our dynamic and uncertain context the suboptimality shows some new features.

Consider the capital stock separated into individual units, and rank them in decreasing order of productivity. When K units are installed, suppose the output of the unit ranked k, for $0 \leq k \leq K$, equals $E(K)\,a(k)$, where E is an increasing function that captures the externality, and a is a decreasing function representing the unit-specific aspects. Now the total output is

$$F(K) = E(K) \int_0^K a(k)\,dk = E(K)\,A(K) \tag{16}$$

say. The marginal product is

$$F'(K) = E(K)\,a(K) + e(K)\,A(K), \tag{17}$$

where $e(K) = E'(K)$ is positive. An increase in K raises output for two reasons: the marginal unit produces its output (the first term on the right hand side), and it raises the output of all previously installed units (the second term).

We want to contrast the process of investment as made by a united entity (large firm or society) that internalizes this externality, and the process when each unit of capacity is controlled by a separate firm or investor who cannot capture the benefit conferred on the others. To make the comparison without confusing it with unrelated issues, some care is necessary. The expected net present value of investment in this model is positive: there are rents to be made. When the right to invest is removed from a united entity, there can be a "rent run" among competing investors. Since this has nothing to do with the externality above, I shall avoid the problem by fixing the order of entry, for example by auctioning the places in a queue of unit-sized firms.[4]

The investor of the marginal unit K captures only the part $E(K)\,a(K)$ of the true marginal product $F'(K)$. Therefore its investment decision proceeds on this basis. Of course this appropriable portion may also show some scale economies; $E(K)\,a(K)$ may be increasing. Then a threshold level of P may trigger a sudden jump of capacity as a mass of firms enters. More properly, this is one possibility, where the jump is expected by all and confirmed by their actions. There may be other equilibria with lock-in at low levels of capital. For example, if $E(0) = 0$, there may be no investment irrespective of P because everyone expects that no one else will enter.

But a particularly interesting case is where private returns are diminishing ($E(K)\,a(K)$ is decreasing) but social returns increasing for a while ($F'(K)$ has an increasing portion). Then the private investment threshold curve is given by

$$P_p(K) = \frac{\beta}{\beta - 1}\,\frac{(\rho - \mu)\,c}{E(K)\,a(K)},$$

[4]For a detailed analysis of rent runs, see Bartolini (1992).

the marginal unit K is installed in equilibrium when this curve is hit.[5] Thus capacity will expand gradually in a decentralized equilibrium, even though a jump is socially optimal.

A simple numerical example shows the nature of the differences vividly. Suppose

$$E(K) = 1 + K, \qquad a(K) = 1 - K.$$

Then

$$E(K)\,a(K) = 1 - K^2, \qquad 0 \leq K < 1,$$

and

$$F(K) = K + \tfrac{1}{2}K^2 - \tfrac{1}{2}K^3, \qquad F'(K) = 1 + K - \tfrac{3}{2}K^2.$$

which makes sense for $0 \leq K < (1+\sqrt{7})/3 \approx 1.21$.

For ease of notation, choose c so that $\beta(\rho - \mu)c/(\beta - 1) = 1$. Then the privately optimal capacity expansion is characterized by the threshold function

$$P_p(K) = 1/(1 - K^2), \qquad 0 \leq K < 1.$$

To find the socially optimal threshold combination, note that the marginal product is maximized at $K^* = 1/3$ with $F'(K^*) = 7/6$, and the right hand branch of the threshold function is given by

$$P_s(K) = 1/(1 + K - \tfrac{3}{2}K^2), \qquad 1/3 < K < (1+\sqrt{7})/3.$$

The left hand branch $Q(X)$ does not have a simple expression, but we can find its general features. The average product is maximized at $K^{**} = \frac{1}{2}$, with $F(K^{**})/K^{**} = F'(K^{**}) = 9/8$, so $Q(0) = P(K^{**}) = 8/9$. Then we can complete the picture shown in Figure 5.

The market equilibrium pattern of investment differs from the socially optimal policy in several respects. First, private firms do not commence investing until the price reaches 1, when the optimal starting price is 8/9. Next, private investment always occurs gradually, while an immediate jump to the capacity level of $K^{**} = \frac{1}{2}$ is socially optimal. Finally, the market will never push capacity beyond 1, while further additions upto about 1.21 are socially optimal if the price rises to sufficiently high but finite levels. The private or appropriable return becomes negative when $K > 1$, but the external benefit keeps the social marginal return positive for a while.

[5] Formal proof needs some care, since the firm installing the marginal unit does so in rational expectation of subsequent investment by other units and a change in its E value. It turns out that this makes no difference to the tradeoff between investing rightaway and waiting another instant; see Leahy (1991).

5 Concluding Comments

I have offered an initial exploration into the theory of irreversible investment when increasing returns and uncertainty are both important. The results are intuitively appealing generalizations of the "real options" approach. More importantly, the analysis should open the way for several further extensions. Some features are in principle easy to incorporate: finitely lived or depreciating capital, large projects requiring a number of stages to be completed, and negative profit flow leading to suspension or abandonment. More substantive modifications include increasing returns over the whole range of capital, as in Romer (1986), and general equilibrium with several sectors. I speculated above that a story of persistent endogenous growth could be told as a stochastic dynamic general equilibrium in which old sectors exhaust their increasing returns and new ones emerge. This has some appeal, but much work remains to be done to complete the picture in a satisfactory way.

Appendix

Here I establish the formulas for the expectations in equations (11) and (12). The first is a standard hitting time formula, for example Harrison (1985, p. 42) or Karlin and Taylor (1975, p. 362), adapted to the context of a geometric Brownian motion. I sketch a derivation using a method more familiar to economists.

Define the expectation in (11) as a function of the starting price P:

$$f(P) = \mathbf{E}\left[e^{-\rho T}\right].$$

So long as $P < Q$, we can choose dt sufficiently small that hitting the level Q in the next short time interval dt is an unlikely event. Then the problem restarts from a new level $(P + dP)$. Therefore we have the recursion relation

$$f(P) = e^{-\rho\, dt}\, \mathbf{E}\, f(P + dP).$$

Expanding the right hand side, recalling that P follows the process (1) and using Itô's Lemma, this becomes

$$f(P) = [1 - \rho\, dt + o(dt)]\, [f(P) + \mu\, P\, f'(P)\, dt + \tfrac{1}{2}\, \sigma^2\, P^2\, f''(P)\, dt + o(dt)].$$

Simplifying and letting $dt \to 0$, we get the differential equation

$$\tfrac{1}{2}\, \sigma^2\, P^2\, f''(P) + \mu\, P\, f'(P) - \rho\, f(P) = 0.$$

This has the general solution

$$f(P) = A\, P^{\alpha} + B\, P^{\beta},$$

where β is the positive root of the quadratic (2), and α is the corresponding negative root.

The constants A and B are determined by a pair of boundary conditions. As P approaches Q, T is likely to be small and $e^{-\rho T}$ close to 1; therefore $f(Q) = 1$. When P is very small, T is likely to be large and $e^{-\rho T}$ close to 0; therefore $f(0) = 0$. Using these, we get $A = 0$ and $B\, Q^{\beta} = 1$, so $f(P) = (P/Q)^{\beta}$ as in (11).

Similarly, define

$$g(P) = \mathbf{E}\left\{ \int_0^T P_t\, e^{-\rho t}\, dt \right\}.$$

This satisfies the differential equation

$$\tfrac{1}{2}\sigma^2 P^2 g''(P) + \mu P g'(P) - \rho g(P) + P = 0,$$

with the general solution

$$g(P) = C P^{\alpha} + D P^{\beta} + P/(\rho - \mu),$$

and boundary conditions $g(Q) = 0$, $g(0) = 0$. Therefore $C = 0$ and $D = -Q^{1-\beta}/(\rho - \mu)$, which gives (12).

References

Bartolini, Leonardo, "Competitive Runs: The Case of a Ceiling on Aggregate Investment," working paper, Research Department, International Monetary Fund, 1992.

Bertola, Giuseppe, "Irreversible Investment," unpublished working paper, Princeton University, 1989.

Brennan, Michael J., and Eduardo S. Schwartz, "Evaluating Natural Resource Investments," *Journal of Business*, Jan. 1985, **58**, 135-157.

Dixit, Avinash, "Entry and Exit Decisions under Uncertainty," *Journal of Political Economy*, June 1989 , **97**, 620 638.

Dixit, Avinash, "Investment and Hysteresis," *Journal of Economic Perspectives*, Winter 1992, forthcoming.

Harrison, J. Michael, *Brownian Motion and Stochastic Flow Systems*, New York: Wiley, 1985.

Harrison, J. Michael and Michael I. Taksar, "Instantaneous Control of Brownian Motion," *Mathematics of Operations Research*, 1983, **8**, 439-453.

Harrison, J. Michael, Thomas M. Sellke and Allison J. Taylor, "Impulse Control of Brownian Motion," *Mathematics of Operations Research*, 1983, **8**, 454-466.

Karlin, Samuel and Howard M. Taylor, *A First Course in Stochastic Processes*, second edition, San Diego, CA: Academic Press, 1975.

Leahy, John, "The Optimality of Myopic Behavior in a Competitive Model of Entry and Exit," Discussion Paper No. 1566, Harvard Institute of Economic Research, August 1991.

Manne, Alan S., "Capacity Expansion and Probabilistic Growth," *Econometrica*, October 1961, **29**, 632 649.

McDonald, Robert, and Daniel R. Siegel, "Investment and the Valuation of Firms When There is an Option to Shut Down," *International Economic Review*, June 1985, **26**, 331-349.

Pindyck, Robert S., "Irreversible Investment, Capacity Choice, and the Value of the Firm," *American Economic Review*, December 1988 , **79**, 969-985.

Pindyck, Robert S., "Irreversibility, Uncertainty, and Investment," *Journal of Economic Literature*, September 1991, **29**, 1110 1152.

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

Skiba, A. K., "Optimal Growth with a Convex-Concave Production Function," *Econometrica*, May 1978, **46**, 527–539.

Weitzman, Martin L. "Optimal Growth with Scale Economies in the Creation of Overhead Capital," *Review of Economic Studies*, October 1970, **37**, 555–570.

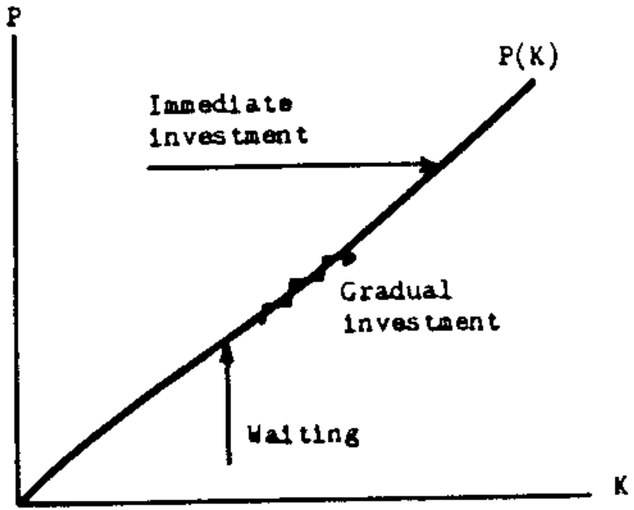

Figure 1 - Optimal Policy without Scale Economies

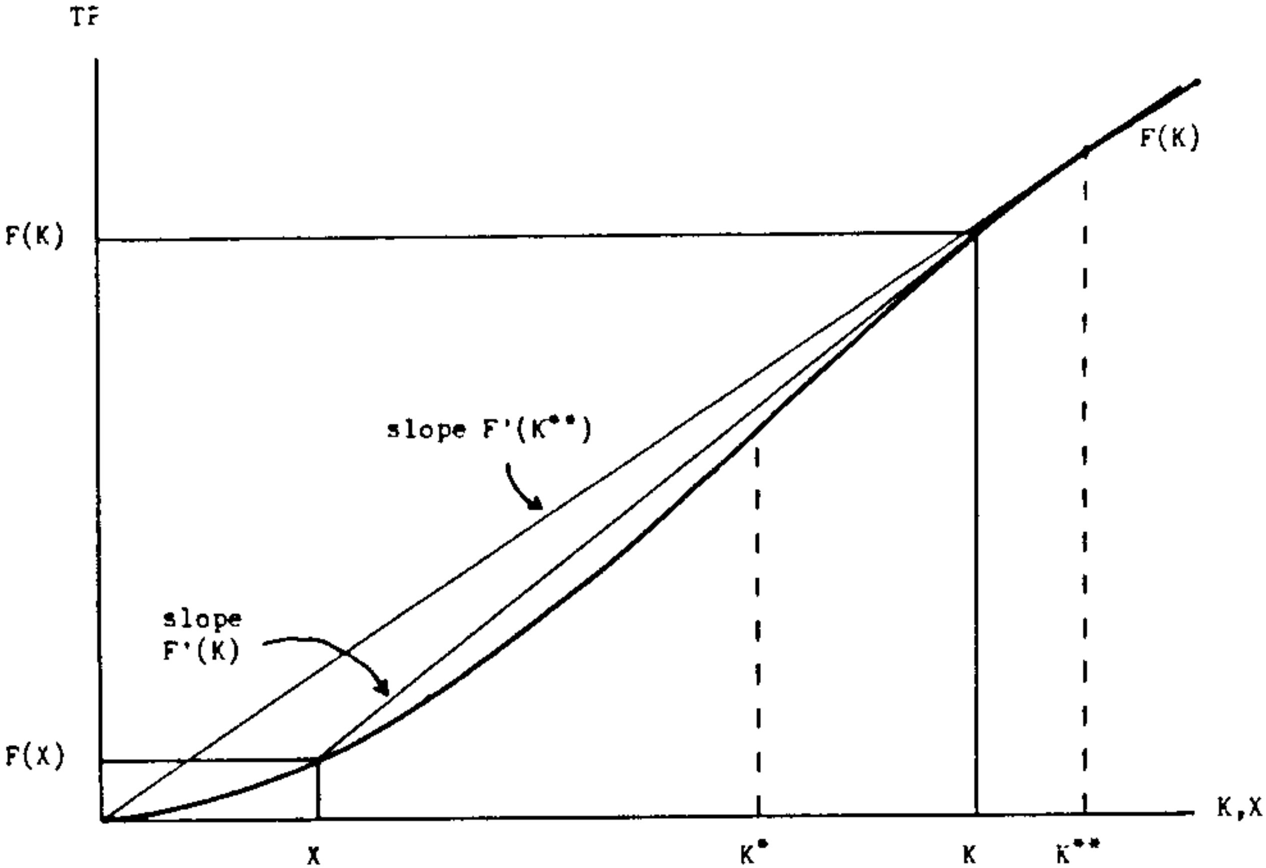

Figure 2 - Total Product

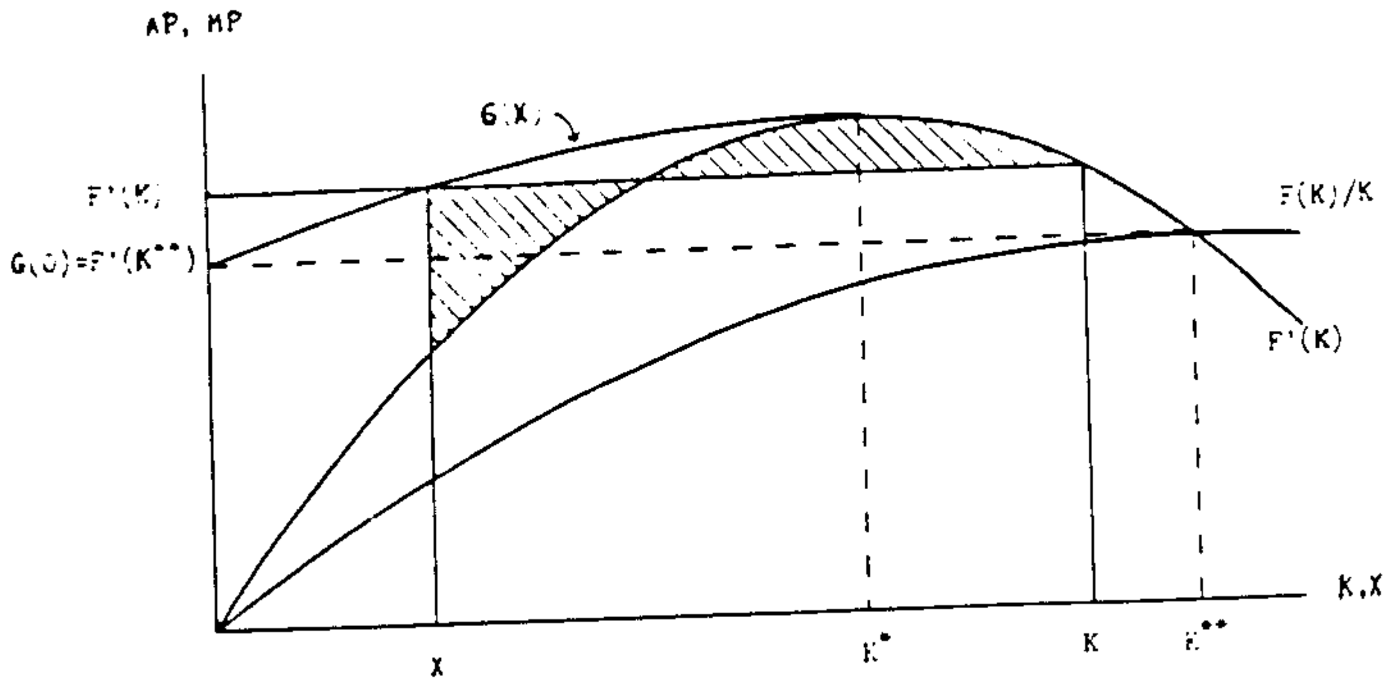

Figure 3 - Average and Marginal Products

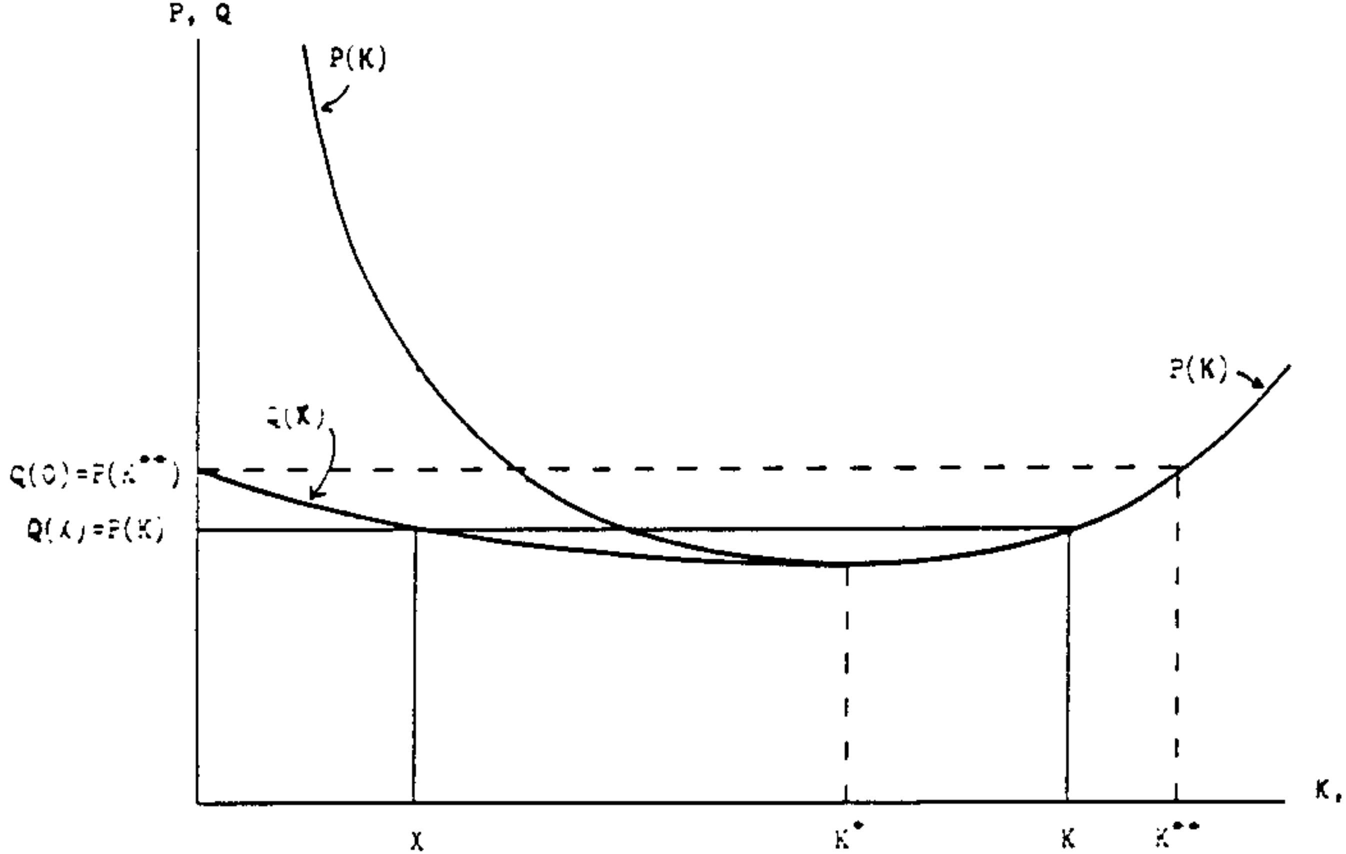

Figure 4 - Optimal Policy with Scale Economies

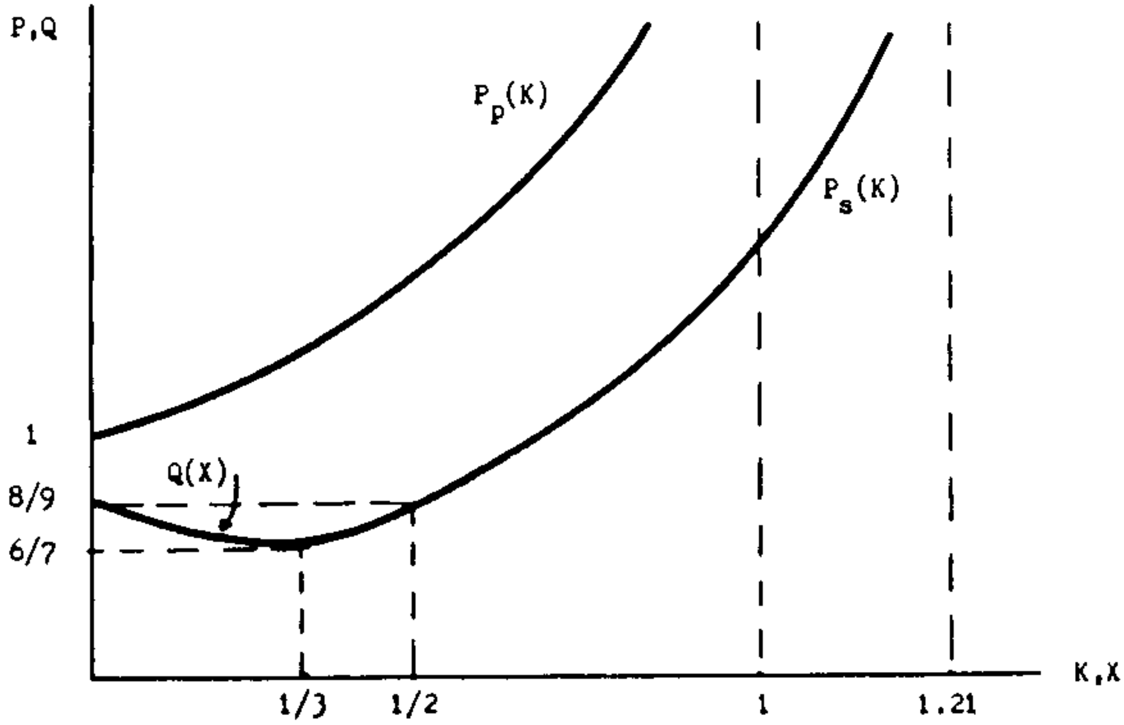

Figure 5 - Example with External Economies