MAIL CODE: EA & EG

WORKING PAPERS

A THEORY OF COST AND INTERMITTENT PRODUCTION:

SOME ANTITRUST IMPLICATIONS

Michael T. Maloney

WORKING PAPER NO. 93

June 1983

BUREAU OF ECONOMICS
FEDERAL TRADE COMMISSION
WASHINGTON, DC 20580

A THEORY OF COST AND INTERMITTENT PRODUCTION:

Some Antitrust Implications

Michael T. Maloney

Clemson University

December 1982: Not for quotation without permission.

I. INTRODUCTION

This paper develops the theory of a firm that has the option of starting and stopping the production process. Intermittent production allows the firm to operate at the cost-minimizing rate regardless of demand conditions. Under certain circumstances costs will be lower and profits higher than if the firm produced continuously. The idea is this, the firm produces at the minimum average cost rate for some period of time during which production exceeds demand and inventories accumulate; then the firm shuts down and sells off its stocks. This paper details the conditions that make this strategy profit maximizing.

In some production situations rate cannot be altered easily. Examples include pipelines, assembly lines, airplanes, and trucks. Driving these machines at rates different from the engineering efficient rate can impose large costs on the firm. However, these costs can be avoided by intermittent production. As an example, transportation firms adjust to slack demand by stopping their vehicles rather than slowing them. Similarly, automobile manufacturers often shut down entire plants while they sell off accumulated stocks.

Factor acquisition and storage costs induce a firm employing intermittent production to diversify its product line until the facility is used continuously. This means that the multiproduct firm is a predictable response to transactions costs. Instead of idlying when the rate of demand is less than the engineering efficient production rate, the firm switches from one good to another producing each at the rate which minimizes average cost.[1]

---

[1]Marketing research has evolved a number of demand side theories of the multiproduct firm. Intermittent production provides an additional motivation for multiproduct diversification--cost savings. For instance, GM has developed an assembly plant that handles three sizes of car frames. Thus,

Multiproduct diversification can be accomplished either through internal expansion or through merger and acquisition. When intermittent production induces firms to expand via merger several predictions emerge. First, the average cost of production will decline. Second, profits will increase. Third, there will be a physical consolidation of capital or whatever resource motivates intermittent production. Fourth, the risk of the financial claims on the company will decline. In this way, the theory of intermittent production can be used as a screen for antitrust prosecution. Other things the same, when two companies merge and both were employing intermittent production before merger, then there is good reason to think that any post-merger increase in profits is in large measure due to lower cost and not increased monoply power. Empirically, intermittent production can be measured by the cyclical nature of inventories. As a first approximation, when two companies merge and both have cyclical inventories before merger and when there is physical consolidation after merger, this theory predicts that merger is more likely for cost than monopoly reasons. Antitrust prosecution should proceed with caution in such cases.

The model can also be used to understand the nature and scope of economies of scale. According to this theory, if economies of scale are ever observed, it is not because of lumpiness and indivisibilities, but rather because there are start-up and shut-down costs, such as capital acquisition costs, uniquely associated with intermittent production. This implies that increasing returns to scale are neither a necessary nor a sufficient condition for economies of scale, and that economies of scale are ultimately a trans-

---

production of its various sized cars can be adjusted without shutting down a plant.

actions cost phenomena.

In the next section, many of the terms employed throughout the paper are defined. Then a zero-transactions cost model is developed which incorporates time and intermittent production. This construction is then used to highlight the relevant transactions costs in the production calculus of a firm which employs intermittent production; a number of predictions are derived. Specifically, Section III demonstrates that economies of scale are solely the result of positive transactions costs, and Section IV shows that product diversification mitigates economies of scale. Section V elaborates on the antitrust implications. A brief conclusion completes the paper.

II. THE PERIOD OF PRODUCTION -- A CHOICE VARIABLE

The role of time in the theory of the firm has a rich background. Contributions include Alchian (1959), Arrow, Karlin, and Scarf (1958), Hirshleifer (1962), Orr (1964), and Hicks (1968). To avoid confusion between this presentation and these earlier papers, let us first define a number of the terms used here. Let $C=C(q,t)$ be the total cost function that maps output levels, q, per time period, t, into dollars, C. Then for any time period t_o and output level q_o, cost is C_o. Initially assume that the instantaneous rate of production is constant over the whole time period so that $dq/dt = q_o/t_o$. Shortly, this assumption will be relaxed and the firm allowed to vary rate as it sees fit, which enables the firm to produce a fraction of time and not produce the remainder. Call that fraction of time the firm operates its period of production and denote it by α, $0 < \alpha \leq 1$.[2] To reiterate, total output over the period divided by time is the average rate of production for the whole period, q_o/t_o, and is equal to the instantaneous rate of production, dq/dt, only if production is continuous and rate does not change.

With these definitions in mind, consider a firm which has a long-run U-shaped marginal cost curve, $C_q(q,t)$[3] for constant and incessant production, embodying first decreasing then increasing costs for reasons outlined as in Robinson (1931).[4] For a given time period, at low levels of output and con-

---

[2]Becker (1971) has addressed the question of the period of production briefly but only as it pertains to a temporary or unexpected increase in demand.

[3]$C_q(q,t) \equiv \partial C(\bullet)/\partial q$.

[4]Robinson discusses the integration of processes, the economies of large machines, massed reserves, large organization, and standardization among other reasons why economies of scale might exist.

stant production, increasing output reduces marginal and average cost as the firm overcomes lumpiness or indivisibilities in production. Each point on $C_q(\bullet)$ represents the marginal cost of that q as the firm produces throughout the entire period. Let $R_q(q,t)$ be the marginal revenue curve for the given time period, t_o. See Figure 1.

The theory assumes that as the time period increases both $R_q(\bullet)$ and $C_q(\bullet)$ shift horizontally by the same proportion as the increase in time. For example, $R_q(q,\ t_o + \beta t_o)$ and $C_q(q,\ t_o + \beta t_o)$ are respectively the marginal revenue and marginal cost curves for $(t_o + \beta t_o)$ time. That is, if the firm can produce q_o output in, say, one month, it can produce $2q_o$ output in two months with no change in marginal cost. This assumption implies that $C(q,t)$ is homogeneous of degree one in q and t.[5] This is not in contradiction with Alchian's proposition that increasing planned volume (holding rate constant) increases cost at a decreasing rate, because the total planned volume of production has not changed. Whatever the planning horizon of the firm, hold it constant. The time period t_o represents a snapshot of the firm, and by increasing the length of time only the exposure period of observation is expanded. Hence, there are no real changes within the firm--rate, total planned volume, and costs are unchanged. All that has happened is to observe the firm longer, for a time period of length $(1 + \beta)t_o$ rather than the originally considered period t_o.

In the usual presentation, Points B and B' in Figure 1 satisfy the first- and second-order conditions for profit maximization in t_o and $(t_o + \beta t_o)$ respectively. When operating in time period t_o, the firm produces q** and

---

[5]One exception to this homogeniety assumption that is well developed in the literature is the analysis of learning curves.

profits are the area A'BA. If this firm were to produce q^* ($=q^{**}(1 + \beta)$) in ($t_o + \beta t_o$) time, profit would be $\Pi(t_o + \beta t_o) = R(q^*, t_o + \beta t_o) - C(q^*, t_o + \beta t_o)$ = A'B'A. In fact, this solution does not maximize profits. Suppose the firm produced quantity q^* in time t_o and sold it in time $t_o + \beta t_o$. Revenues would still be OA'B'q*, but costs would be reduced to OADq*. Profits would increase by AB'DA. Ultimately, profits are not maximized unless the firm produces at the rate of minimum average cost.

This result is developed next, but first let us summarize the structure of the model. Implicit in the preceding discussion are the assumptions:

(1) Both cost and revenue are homogeneous of degree one in output and time,
(2) all resources are costlessly liquidated at the end of the production period,
(3) the firm can store output costlessly,
(4) there are no extra costs associated with start-up and shut-down, and
(5) the discount rate is zero.

Since the total cost function is homogeneous of degree one in output and time, it is easily shown that $C_q(\bullet)$ and $C(\bullet)/q$ are both homogeneous of degree zero in q and the period of production, α. Consequently, the marginal and average cost curves for constant, incessant production are compressed in a horizontal fashion for reductions in the period of production, and a family of cost curves exist for $0 < \alpha \leq 1$. Each α determines an average cost curve for uninterrupted continuous production over the period αt_o, during which average cost changes with the instantaneous rate of output. From the family of average cost curves an envelope exists that is horizontal, as it is the minima of these average cost curves. In Figure 2, three such average cost curves are shown where α takes the values α', α'', and 1. The interval (A, B], is the minimum average cost envelope for all values of α between zero and one, and therefore it is the firm's actual long-run average cost curve over this range of output. This cost function is identified by the three variables q, α, and

t_o: $C = C(q, \alpha, t_o)$.

The profit function for the firm is:

$$\Pi(t_o) = R(q, t_o) - C(q, \alpha, t_o) \quad .$$

Maximizing profit with respect to output and the period of production, subject to the constraint $\alpha \leq 1$, yields:

$$(1) \qquad R_q(\bullet) = C_q(\bullet) \qquad \text{and}$$

$$(2) \qquad -C_\alpha(\bullet) = \kappa,$$

where κ is the Kuhn-Tucker multiplier on the constraint. If the constraint is binding, the firm produces continuously and κ is the marginal effect on profits of not being able to increase total output without increasing the rate of production. In this case, equation (1) says that the firm responds to increases in marginal revenue by moving up the marginal cost curve and increasing the rate of production. If the constraint is not binding, $\kappa = 0$ and the sufficient, second-order conditions are:

$$(3) \qquad -C_{\alpha\alpha}(\bullet) < 0 \qquad \text{and}$$

$$(4) \qquad -C^2_{\alpha q}(\bullet) < C_{\alpha\alpha}(\bullet)[R_{qq}(\bullet) - C_{qq}(\bullet)] \quad .$$

Conditions (2) and (3) now say that the choice of α must minimize total cost at the optimal level of output. This is equivalent to minimizing average cost. When $\alpha < 1$, variations in q are accompanied by changes in α. By the five assumptions stated above, the cost of expanding simultaneously at the time and output margins is constant and equal to minimum average cost.

This result is depicted graphically in Figure 2. Marginal revenue equals marginal cost at q' units of output. If the production period is α', then the average cost of production is E. With a production period of any other length, the average cost of q will be higher. For example, with a shorter production period, α'', average cost is G. If the firm produces for the entire period, it has an average cost of H; both G and H are greater than E. That is, as α is decreased from α' to α'' or increased to 1, average cost increases. Thus, profit is maximized by choosing a period of production so that marginal revenue is equal to minimum average cost. In Figure 2, only $\{q', \alpha'\}$ satisfies conditions (1) - (4).

The important result is that, in the zero transactions-cost case, the minimum average cost envelope characterizes the behavior of the firm. As marginal revenue changes, profit-maximizing output adjusts along this locus. This means that not only does a monopolist operate at the minimum average cost during the period of production, but during its production run, the monopoly firm also produces at the same instantaneous rate as a (comparable) perfectly competitive firm.

Suppose the firm experiences a permanent decrease in marginal revenue. Conventional wisdom tells us that a monopolist will decrease its production. This is not exactly correct. Total volume over any specified time period will of course decrease, but not the instantaneous rate of production. The firm will choose a shorter period of production to accommodate the decline in sales, but while it produces, its rate of production is unchanged by the decline in marginal revenue.

To increase the empirical relevance of the theory, it is useful to relax the simplifying assumptions of this model. Discounting becomes important since costs and revenues are not paired. Production costs are borne only during the period of production, while revenues are received continuously over t_o. Consequently, an increase in the interest rate induces the firm to increase the period of production. As the firm stands at time zero and plans production over time t_o, an increase in the discount rate will force the firm to forego some current production (costs) whose present discounted value is relatively lower. Hicks (1968, pp. 213-217) concludes much the same in his treatment of interest and production. The problem is treated more carefully in an appendix.

Storage, Start-up and Shut-down Costs

Storage costs undermine the firm's incentive to employ intermittent production. When these costs are important, the firm reacts by choosing numerous shut downs of short duration, which drives storage costs toward zero. Start-up and shut-down costs hinder this reaction and thereby decrease the value of intermittent production.

A convenient method of including storage costs is to assume that the firm buys storage from a perfectly competitive, constant-cost storage industry. Storage adds a cost per unit stored, per unit of time stored, to the optimization problem. Inventory accumulates at the rate of $q/\alpha t_o - q/t_o$ per instant of time while production proceeds, and at the rate of $-q/t_o$ when production is shut-down or interrupted. Figure 3 depicts the inventory accumulation-depletion process. The height of I at any t gives the stock of inventory at that time for a particular output level and period of production. Storage costs will be a function of the area under I.
Inventories are

$$I(t) = \begin{cases} (q/\alpha t_o - q/t_o)t & t \leq \alpha t_o \\ q - (q/t_o)t & t \geq \alpha t_o \end{cases}$$

for any q and α, and storage costs, S(t), are

$$S(t) = k \int_0^t I(u)\, du,$$

where k is the per unit per instant time storage fee.

When the firm produces only during the first portion of t_o, its storage bill will be $kq(1-\alpha)t_o/2$. However, the firm can reduce its storage bill by breaking up the production period. As shown in Figure 3 by the dashed lines, the area under the inventory path is smaller when production is accomplished by two runs of $\alpha t_o/2$ instead of one of αt_o. In fact, the firm can break the process into many production periods, n, each of length $\alpha t_o/n$ accompanied by a shut-down of $(1-\alpha)t_o/n$. When the firm uses multiple start-ups its inventories for each of these start-up-shut-down cycles are:

$$I(t) = \begin{cases} (q/\alpha t_o - q/t_o)t & t \leq \alpha t_o/n \\ q/n - (q/t_o)t & t \geq \alpha t_o/n \end{cases} .$$

Storage costs are $kq(1-\alpha)t_o/2n$ and diminish to zero as the number of production breaks goes to infinity. However, starting and stopping production will add costs: label these S_1 and S_2. The optimal number of start-ups and shut-downs occurs when the reduction in storage costs is equal to the start-up plus shut-down cost, that is, where

$$(5) \quad kq(1-\alpha)t_o/2n^2 = S_1 + S_2.$$ [6]

Since storage costs approach zero as production is divided into many short periods, the average cost curve of the firm remains flat even when storage is incorporated into the problem so long as start-up and shut-down costs are zero. Also, as the firm breaks the time period t_o into many short production periods, each followed by sales with no production, the importance of positive interest rates is mitigated. When there are a large number of production periods during the relevant period t_o, production and sales are nearly paired, and the firm will not be as concerned about discounting the future

[6]The term $[-kq(1-\alpha)t_o/2n(n-1)]$ is the difference in storage costs for integer reductions in n. In the limit as $\Delta n \to 0$, $\partial S/\partial n = -kq(1-\alpha)t_o/2n^2$.

revenues because they are received almost simultaneously with costs.

The first order conditons for profit maximization identified in equations (1) and (2) have to be modified to include equation (5) when the firm is allowed to choose the number of start ups over the production period. Given that the firm can choose q, α, and n then the sufficient second order conditons require that the hessian of second partials be negative definite.

$$H = \begin{matrix} R_{qq} - C_{qq} - S_{qq} & -C_{q\alpha} - S_{q\alpha} & -S_{qn} \\ -C_{\alpha q} - S_{\alpha q} & -C_{\alpha\alpha} - S_{\alpha\alpha} & -S_{n\alpha} \\ -S_{nq} & -S_{n\alpha} & -S_{nn} \end{matrix} .$$

There are two sets of implications that are interesting, those resulting from changes in per unit storage costs, k, and those from changes in start-up and shut-down costs, $(S_1 + S_2)$. Based on the sufficient second order conditions, one can predict that the number of production intervals decreases with an increase in start-up and shut-down costs. That is, $dn/d(S_1 + S_2)$ is negative because this derivative is equal to the ratio of the second and third order principle minors of the hessian of second partials shown above.[7] The problem assumes these alternate in sign. This supports the inference above that storage costs are ultimately a product of start-up and shut-down costs.

The model cannot predict what will happen to output and the length of the production period when per unit storage, start-up and shut-down costs change as the signs of $dq/d(S_1 + S_2)$, $d\alpha/d(S_1 + S_2)$, dq/dk, $d\alpha/dk$, and dn/dk derived from the hessian H are all indeterminant. Returning to Figure 2, some inferences concerning the effect of storage costs on output can be drawn, and then

---

[7] Third order minor means the full hessian.

the analysis can be extended by looking at the choices of the firm when minimizing cost subject to an output constraint.

Consider the discrete change in the storage problem where storage costs go from zero to infinity. When these costs are zero, the firm produces where marginal revenue is equal to minimum average cost by appropriately adjusting the period of production. When storage costs are prohibitive, the firm produces continuously where marginal revenue is equal to the marginal cost function given by $\alpha = 1$. Call the former case A and the latter case B. It is immediately apparent that when marginal revenue intersects the marginal cost for case B above the minimum average cost locus of case A, a discrete decrease in storage costs from infinity to zero will cause output to increase. On the other hand, if marginal revenue intersects marginal cost for case B below the minimum average cost locus of case A, the decrease in storage costs will cause output to fall. The effect on α is the same in both cases, the change in α for a change in S from infinity to zero is positive. This is because the argument deals with the polar extremes of the storage problem. For changes of smaller magnitudes, and for changes in k and $(S_1 + S_2)$ independently, the effect on both q and α is indeterminant. Even so, the analysis is revealing. For reductions in storage cost at low output levels, output will tend to increase, whereas at larger levels this prediction is reversed. Low and high output levels are defined in terms of how fully the firm is taking advantage of technical economies of scale. Because the output effect can go either way it makes sense to examine the behavior of the firm holding output constant.

The comparative static results of minimizing cost subject to an output constraint can be reviewed by reference to the following equations derived from the sufficient, second order conditions:

(6) $d\alpha/dk = q^2 t_o^2 n^{-4} k(1-\alpha)/4H > 0$

(7) $d\alpha/d(S_1 + S_2) = S_{\alpha n}/H > 0$

(8) $dn/dk = [C_{\alpha\alpha}(1-\alpha) - (1/2)kqt_o n^{-1}]/[2C_{\alpha\alpha}(1-\alpha) - (1/2)k^2 qt_o n^{-2} > 0,$

if $k > 2n$, and

(9) $dn/D(S_1 + S_2) = - C_{\alpha\alpha}/H < 0.$

These results square with our intuition. From equations 6 and 7, the higher the cost of storage either because of a larger per unit charge or a larger cost of starting and stopping, the longer the fraction of t_o the firm spends in production. Similarly from equation 8, the more costly the per unit charge, the more the firm will attempt to mitigate storage by starting and stopping production, thereby reducing the amount of inventories as shown by the dashed line in Figure 3. There may be an upper bound on this strategy, however.[8] Finally, the relationship of n to $(S_1 + S_2)$ is unchanged from the profit maximizing model. The conclusion of this exercise is that start-up and shut-down costs are the driving factor in determining the firm's use of intermittent production as a cost reducing strategy.

Summary

The period of production is a technique that isolates the the cost-minimizing rate of production from the vagaries of demand. To accommodate its customers, the firm expands or contracts the length of its production process and so enjoys whatever is the cheapest way to produce output. Consequently, in the case of zero start-up/shut-down costs, demand only determines the period of production and the volume of production, not the production rate.

---

[8]For the denominator of the expression for dn/dk to be positive, as it must if the sufficient second order conditions are satisfied, it is necessary that $C_{\alpha\alpha}(1-\alpha) > k^2 qt_o n^{-2}/4$, in which case, the whole expression is positive if $k > 2n$. If $k < 2n$, the sign of dn/dk depends on the magnitude of $C_{\alpha\alpha}(1-\alpha)$ relative to k, q, and n.

Start-up and shut-down costs most likely result from the acquisition and liquidation of resources. Recall that when production is halted, all factors are liquidated and must be reacquired when production resumes. To the extent that factor liquidation and acquisition costs are not zero, the value of the period of production as a cost-minimizing technique is mitigated. This point is the focus of the remaining sections.

III. ECONOMIES OF SCALE

It should be apparent from the preceding discussion that indivisibilities and lumpiness do not provide an empirically relevant explanation for economies of scale. No matter what the physical relationships in production are, when the costs of intermittent production are zero, there are no economies of scale. In a zero transactions cost world, firms can costlessly acquire capital, either by rental or by purchase and then liquidate it. Labor and management can be purchased at their respective market wages regardless of the contract length. In other words, resource prices do not vary with the number of hours contracted. Hence, every production technology is available to any firm to produce any desired output level at the same price. Quite naturally in a frictionless world, there are no economies of scale, but, though that world contains no frictions, it does contain some fictions. Most contracts require initializing negotiations that are sunk costs at the actual time of delivery. Similarly, when labor must be paid more per hour for one hour of work than for a forty-hour week, or when it is not free to travel to the capital rental store, the average-cost curve need no longer be perfectly elastic.[9]

Consider Figure 4 where for comparison purposes three cost curves are drawn. AC_0 is the frictionless intermittent production cost curve. AC_1 is the uninterrupted production average cost curve that embodies economies of scale because of lumpiness and indivisibilities. It is the upper limit of the firm's actual average cost function. If start-up and shut-down costs are so

---

[9] In the construction trades, quite often it is virtually free to travel to the capital rental store because the machinery must be moved to the construction site anyway. Thus, firms are more likely to rent such equipment than to own it except when there are large agency costs of operation such as described in the parable of the hammer [Alchian and Demsetz (1972)]. Firms whose production is mobile are more likely to rent than own their capital, *ceteris paribus*.

large that the firm must produce constantly, then cost curve AC_1 describes the firm's activity and is the only case where economies of scale exist because of increasing returns in production. In all other circumstances, AC_2, the synthesis of AC_o and AC_1, embodies costly intermittent production and characterizes the firm.[10] Average cost declines in AC_2 because the firm is unable to fully enjoy returns to scale due to start-up and shut-down costs. In the limit, as start-up and shut-down costs approach zero and as the number of start-ups goes to infinity, the cost curve approaches AC_o. As start-up costs increase, the optimal number of interruptions diminishes toward zero, and consequently the cost curve rises, approaches AC_1, and equals AC_1 when $n = 1$.

This means that economies of scale are an economic not a technological phenomena. Costly start-ups and shut-downs lift the left end of the cost curve, and so the elasticity of average cost is only a function of these costs. In other words, start-up and shut-down costs are necessary and sufficient for economies of scale; increasing returns are neither.

More than simply demonstrating that cost curves decline because of an unusual kind of factor rigidity, analysis of the period of production implies that output adjustments will usually be associated with shut-downs as well as, if not instead of production rate reductions. The curve AC_2 implies that the firm is shut down some of the time. Intermittent production, although costly, is less expensive than the reduced instantaneous production rate necessary to accommodate a continuous output flow.[11] The importance of this result is that

---

[10]We know that intermittent production is costly based on the comparative static results of the last section. If either or both of the storage cost components fall, the firm will produce more intermittently.

[11]This suggests the basis for a real business cycle. Each firm operates cyclically. The aggregation of firms may also operate cyclically. Temporary periods of mass shut-downs (unemployment) may be a predictable macroeconomic response to cost-minimizing intermittent production.

one can expect to see firms adjusting their period of production as well as their output rate when demand fluctuates.

IV. THE MULTI-PRODUCT FIRM

Proctor and Gamble produces a number of soap products, Tide, Cheer, and Oxydol among others. But they do not produce each soap simultaneously at any one plant. The production facility is used to produce one soap and then another intermittently as described in the preceding analysis. For example, Tide is run for several days at a rate in excess of sales and stored for future sales. Another soap is then produced at the same instantaneous rate as Tide (although for a shorter period of production since Tide is the big seller). As predicted, if demand grows for any one of these soaps, the instantaneous rate of production is not increased, rather the period of production is extended to accommodate the increase in sales.[12]

Proctor and Gamble produces several soaps for obvious reasons. Given input prices, there is a specific production rate which minimizes cost. Apparently, this rate exceeds the rate at which any one soap is demanded. Rather than producing a number of brands, suppose that P-G only produced Tide, and the other soaps were produced by their own self-contained firms. P-G could produce Tide in the same fashion as it does now, and then lease or sell its equipment to the producer of Cheer. This is the frictionless story described earlier. However such an exchange is not free, hence the merger of Tide and Cheer. Labor and capital can be purchased to produce Cheer at a weekly rate even though the production run may be only a few hours, because the inputs are converted from the production of one soap to another. One can expect the merger of Tide with other soaps until the production facility is

[12]If the facility is used full time, demand fluctuations must be accommodated by adjustment in the period of production of two brands. In the case of Tide, P-G produces a store brand soap the supply of which is adjusted as the demand for the other brands dictates. This example was provided to us by Dan Orr, a former Procter and Gamble employee.

used full time. This merger, at one plant, makes the transfer of resources associated with intermittent production less expensive.[13]

The characterization of this phenomenon in terms of a single plant results from our focus on capital as the input most costly to start up and shut down. However, other resources such as technical personnel may also qualify for period of production consideration. Gort (1962) provides empirical support for the hypotheses that big plants and firms with a large proportion of technical personnel are most diversified.

Thus far, the discussion has centered on long-run decisions under the condition of stable demand. In this case, a firm with downward sloping demand is likely to produce intermittently and produce more than one good. If randomly fluctuating demand and U-shaped average variable cost curves are considered, the same behavior is predicted for competitive firms facing flat demand curves. The traditional treatment of competition shows firms choosing plant sizes based on minimum average cost and accommodating short-run demand fluctuations by varying production rate along the associated short-run marginal cost curve. The period of production analysis predicts adjustment in another dimension--the length of production runs. By varying the period of production, the firm producing at the rate of minimum average variable cost flattens its cost function. If demand declines sufficiently, one expects to see competitive firms ceasing production from time to time, and selling off stocks. Halting production of one item may not mean idling the facility, however, as the product diversification argument reveals. For instance, when the price of

[13]Pfouts's (1961) theory of the multiproduct firm proceeds along similar lines although he generally takes as given the variety of products being produced. In contrast, the discussion here is driven by a desire to explain the origin of the multiproduct firm and which products will be produced under a single firm rather than separately.

wheat falls, farmers transfer production to cattle, hay, or soybeans. They do not choose a slower-growing variety of wheat.

There may be extra organizational costs associated with multiproduct firms, in which case the firm is expected to add products until the extra organizational costs equal the marginal cost savings from fewer start ups. Intuition suggests that these start-up costs will be lowest within the firm when the productive resources have to undergo the slightest physical change. It is reasonable then to expect mergers of firms that produce different models of cars, many pieces of furniture, bolts of different length, screws with metric and American threads, paints with different bases, and the like.[14]

One can think of many reasons why a firm would produce several products. The cost savings from employing intermittent production provide an additional motivation for the firm to undertake product expansion. This is another version of the Coasian firm where transactions within the firm are cheaper than market transactions. The model of intermittent production suggests that the ability to mitigate returns to scale derives from adjusting the period of production and from diversification into different markets between which the transfer of resources is possible. The diversified or multiproduct firm is basically a device for reducing the transaction costs of buying resources on a short-term basis, using them intensively in order to achieve minimum average cost, and then selling them to another who has the same intent. This argument parallels the theory advanced by Stigler (1951) as well as Coase (1937). When

[14]The firm may also expand its product line to achieve full utilization of consumer brand-name recognition. The plant size is then chosen to satisfy demand. The brand-name explanation does not explain why the firm produces many products in each of many plants, however. Even so, brand-name-resource utilization is entirely consistent with the theory developed here. If the inputs necessary to create and maintain brand-name quality are not fully utilized, expansion of the product line is an obvious solution.

the division of labor is limited by the extent of the market, firms will expand into new markets, any market requiring similar resources. They do this to reduce the transactions costs of buying and selling resouces that can only be used efficiently in an intensive manner.

V. MERGERS, INTERMITTENT PRODUCTION, AND ANTITRUST POLICY

A cost-based theory of mergers has been developed in at least two places: the property rights theory of the firm and the modern theory of finance. Both are closely associated with the theory of intermittent production and cost. Finance research has argued that mergers can occur based on three motives: to monopolize, to purge inefficient management, and to enjoy synergistic gains. Obviously the notion of synergy is a catch-all catagory, but the theoretical efforts and empirical results have attempted to separate managerial malfeasance from optimal control structure within the corporation. Bradley's (1980) empirical results support the idea that firms engage in tender offer takeovers for reasons other than replacement of management. At the same time, the monopoly motive is hard to reconcile with these results inasmuch as firms targeted by unsuccessful tender offers usually experience value increases. Thus, the finance literature has a set of empirical results in search of a theory.

The property rights or organizational cost theory of the firm, especially recent contributions by Alchian and Demsetz (1972) and Klein, Crawford, and Alchian (1978), sketch a general picture of the synergistic merger motives popular in finance. Contract enforcement is the key determinant of organizational structure. K-C-A argue that integration of productive activities within the boundary of one firm may be the least cost mechanism of preventing one or the other party to a contract from opportunistically reneging on the contract terms. Why integration might work sometimes and not others is unexplained. Barzel (1981) adds to this theory by suggesting that some transactions must take place at fixed prices in order to prevent post-contractual opportunistic behavior. Again, this is accomplished by organizing production under the umbrella of the firm.

The theory of intermittent production builds on the motivation for merger due to organizational costs. Undiversified firms using intermittent production will usually have resources for which the cost of liquidation and reacquisition limit period of production adjustments. The cost savings to be had from producing intensely are gained by expanding the product line. This expansion can come either through internal growth or merger. In all events, if the period of production is not fully employed as a cost minimizing device, it must mean that high contracting costs make diversification, integration, and possibly merger a relatively attractive strategy.

This theory of cost adds to the contracting-cost-merger discussion in an empirically important way: The firms which are likely candidates for merger based on this motive have distinguishing characteristics. Firms that have cyclical inventory patterns, and especially cyclically idle storage capacity, are definitionally encumbered by high transactions costs. Thus, they are prime candidates for diversification and merger. Merger should occur with other firms similarly situated, thus reducing the idle time of the facilities of both enterprises. Cyclical inventories in each line will persist, but overall capacity utilization is increased. This will be accomplished via physical consolidation.

The comparative static results of section II are indecisive concerning the effect on total inventories. Thus, one cannot make a prediction concerning the amplitude of the pre- and post-merger inventory cycles. Merger causes start-up and shut-down costs to fall. The likely effect of the decline in these costs on output is positive and on the period of production, negative. The predictable effect on the number of shut downs is positive. Maximum inventories, which are a measure of the amplitude, are $q(1-\alpha)/n$, so the effect of the merger is unknown. In all events, inventories will not decline to zero, though they may decline in all product lines. Hence, pre- and post-merger inventory cycles, not necessarily of the same shape, are a signal that the merger is likely to be one based on cost motives rather than monopolization.

Another empirically relevant prediction is that a consolidation of productive facilities should accompany the merger when machines and physical capital are the input motivating intermittent production. Of course, other inputs can qualify for period of production consideration. For example, one firm might have a computer software package that directs production, allocates inputs, and dispatches output. Suppose the computer package is only used part-time. The firm can either sell those services or expand its product line to increase utilization of the computer center. Suppose that outside sale would necessarily reveal trade secrets, such as the computer program itself. Then the firm will have to expand. This merger or acquisition of new product lines will not have to lead to physical consolidation of production facilities, but instead consolidation of the management faculties.

Finally, a rather subtle financial market prediction emerges from the theory, which is that the beta coefficient of the post-merger firm should be lower than the weighted average of the pre-merger companies. This prediction is consistent with common knowledge concerning all mergers. However, common knowledge has been shown to be at odds with the modern theory of finance. Intermittent production offers revived theoretical foundation.

From the demand side of the market alone, the modern theory of finance argues that firms will not merge in order to achieve lower risk. Firms undertake all positive net present value projects regardless of the outcome on their relative riskiness, letting investors diversify through the more efficient mechanism of the stock market.

The theory of intermittent production argues that firms merge in order to enjoy lower costs. However, as a free by-product, the diversity over product lines gives them a wider range of productive opportunities when demand fluctuates. From this comes the prediction on the beta coefficient of the merged firm. Mandelker (1974) finds, for mergers of all types between 1941 and 1962, that post-merger betas are approximately 8% lower than pre-merger betas but not in a statistically reliable sense. A sub-sample of his firms that have cyclical production before merger should show a significant difference in the pre-merger and post-merger betas.

As Peltzman says,

> ... a causal relationship running from concentration to profitability can operate either through an effect on price ..., or on average cost, or, of course, both.
> (1977, p. 229)

The theory presented here develops techniques to help differentiate cost based profitability form monopoly induced profitability. As such, it may prove useful in antitrust case selection. Failure to recognize these principles can lead antitrust authorities to prosecute cases where there are little or no benefits.

VI. CONCLUSION

This paper develops a theory that models the decision calculus of the firm by incorporating the period of production as a choice variable. In a zero transactions-cost world, the firm can produce at the rate which yields minimum average cost, regardless of demand conditions, by varying the length of the production run. However, start-up and shut-down costs reduce the cost savings from this strategy. This means that economies of scale are a transactions cost, not a technological phenomena.

In some cases, firms find it profitable to simply idle their facilities from time to time while selling off accumulated stocks. In other cases, start-up and shut-down costs are sufficient to prevent this direct period of production adjustment as a cost minimizing technique. However, multiproduct diversification is an indirect method for firms to enjoy minimum average cost production rates where demand is insufficient to achieve that end. The start-up and shut-down costs associated with intermittent production are expected to stem largely from the transactions costs of acquiring and liquidating resources, and there are products for which these transactions are most cheaply accomplished through the mechanism of a firm. This multiproduct diversification, usually within a single plant, allows the firm to ignore demand when it chooses its rate of production and operate at whatever rate is the cheapest. In summary, multiproduct diversification increases the elasticity of average cost for each product the firm manufactures.

References

A. A. Alchian, "Costs and Outputs," in The Allocation of Economic Resources by Moses Abromovitz and others, Stanford: Stanford University Press, 1959, 23-40; reprinted in Readings in Microeconomics, 2nd edition by W. Breit and H. M. Hochman, Hinsdale: Dryden Press, 1971.

_______ and H. Demsetz, "Production, Information Costs, and Economic Organization," American Economic Review, LXII (5) (December 1972): 777-795.

K. J. Arrow, S. Karlin, and H. Scarf, Studies in the Mathematical Theory of Inventory and Production, Stanford: Stanford University Press, 1958.

Yoram Barzel, "The Firm: A Coordinator of Contracts," working paper, U. of Washington, 1981.

G. S. Becker, Economic Theory, New York: A. A. Knopf, 1971.

Michael Bradley, "Interfirm Tender Offers and the Market for Corporate Control," J. of Business 53 (4) (October 1980): 345-376.

R. Coase, "The Nature of the Firm," Economica, IV (1937): 386-405.

M. Gort, Diversification and Integration in American Industry, National Bureau of Economic Research, Princeton: Princeton University Press, 1962.

J. R. Hicks, Value and Capital, 2nd edition, Oxford: Oxford University Press, 1968.

J. Hirshleifer, "The Firm's Cost Function: A Successful Reconstruction." Journal of Business, XXXV (3) (July 1962): 235-255.

B. Klein, R. Crawford, and A. Alchian, "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," J. of Law and Economics XXI(2) (October 1978): 297-326.

G. Mandelker, "Risk and Return: The Case of Merging Firms," Journal of Financial Economics, 1 (1974): 303-335.

D. Orr, "Costs and Outputs: An Appraisal of Dynamic Aspects," Journal of Business, 37 (January 1964): 51-60.

Sam Peltzman, "The Gains and Losses from Industrial Concentration," J. of Law and Economics XX(2) (October 1977): 229-264.

R. W. Pfouts, "The Theory of Cost and Production in the Multi- Product Firm," Econometrica, 29 (4) (October 1961): 650-658.

E. A. G. Robinson, The Structure of Competitive Industry, Cambridge Economic Handbooks, Chicago: The University of Chicago Press, 1931.

G. J. Stigler, "The Division of Labor is Limited by the Extent of the Market," Journal of Political Economy, (June 1951): 185-193.

APPENDIX

If continuous discounting of costs and revenues is introduced because the two are not paired, then the present value to the firm of the production process over time t_o is given by:

$$\text{(i)} \quad W = \int_0^{t_o} [R(q,t_o)/t_o]\, e^{-ru}\, du - \int_0^{\alpha t_o} [C(q,\alpha,t_o)\;/\alpha t_o]\, e^{-ru}\, du$$

because the instantaneous revenues and costs are $R(q,t_o)$ and $C(q,\alpha,t_o)$ respectively. Integation of (i) yields:

$$\text{(ii)} \quad W = [R(q,t_o)/t_o]\bullet[(1-e^{-rt_o})/r] - [C(q,\alpha,t_o)/\alpha t_o] \bullet[(1-e^{-r\alpha t_o})/r]$$

The first-order conditions necessary to maximize (ii) with respect to q and α are given by:

$$\text{(iii)} \quad [(1-e^{-rt_o})/rt_o]\, R_q(\bullet) = [(1-e^{-r\alpha t_o})/r\alpha t_o]\, C_q(\bullet)\ ,\ \text{and}$$

$$\text{(iv)} \quad (1/\alpha)[(1/\alpha t_o)C(\bullet)(1-e^{-r\alpha t_o})(1/r) - (1/t_o)C_\alpha(\bullet)(1-e^{-r\alpha t_o})(1/r) - C(\bullet)e^{-r\alpha t_o}] = 0.$$

(iv) is the new first-order-condition corollary to $-C_\alpha(\bullet) = 0$. From (iv) it is easily shown that $C_\alpha(\bullet)$ is now positive. Thus, the optimal period of production is greater with discounting (positive interest rates) than without. For simplicity, all production has been placed at the beginning (first αt_o) of the production period. As noted in Section III, this is not necessarily cost minimizing. To the extent that there is more than one shut-down in the time period, the discounting role becomes less important as costs and revenues become more nearly paired.

Figure 1

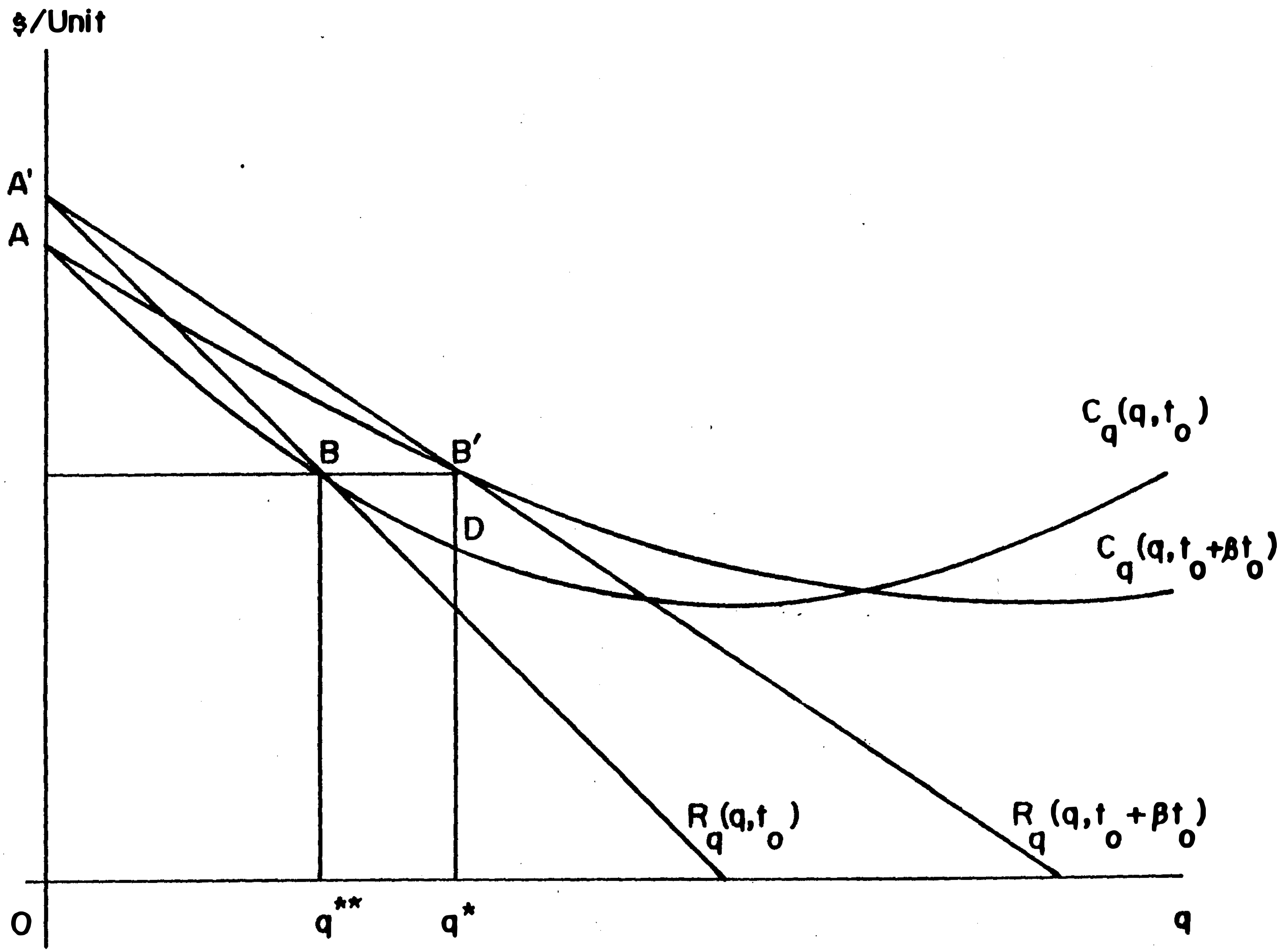
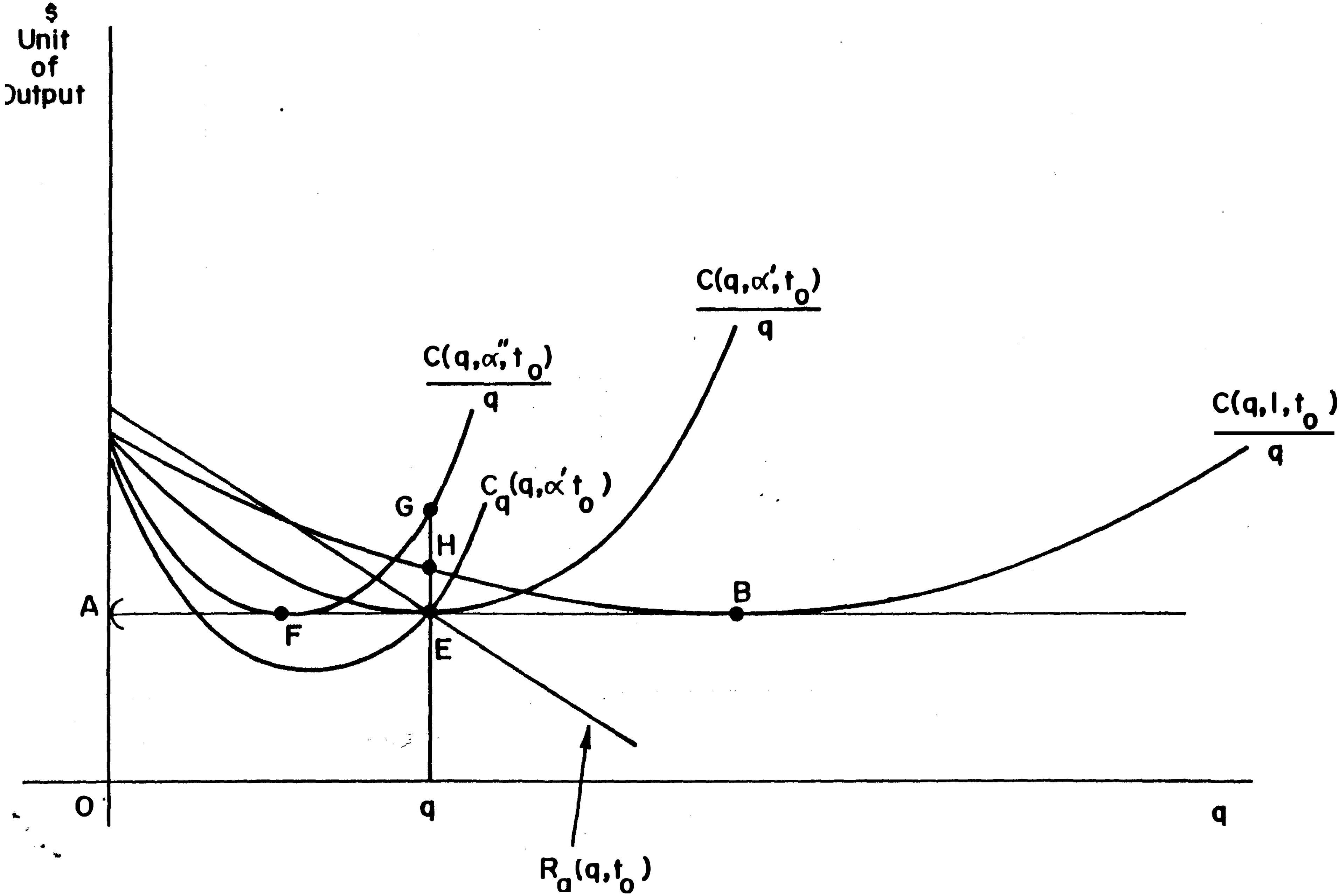


Figure 2



$
Unit
of
Output
C(q,α″,t₀)/q
C(q,α′,t₀)/q
C(q,1,t₀)/q
Cq(q,α′t₀)
G
H
A
F
E
B
O
q
q
Ra(q,t₀)

Figure 3

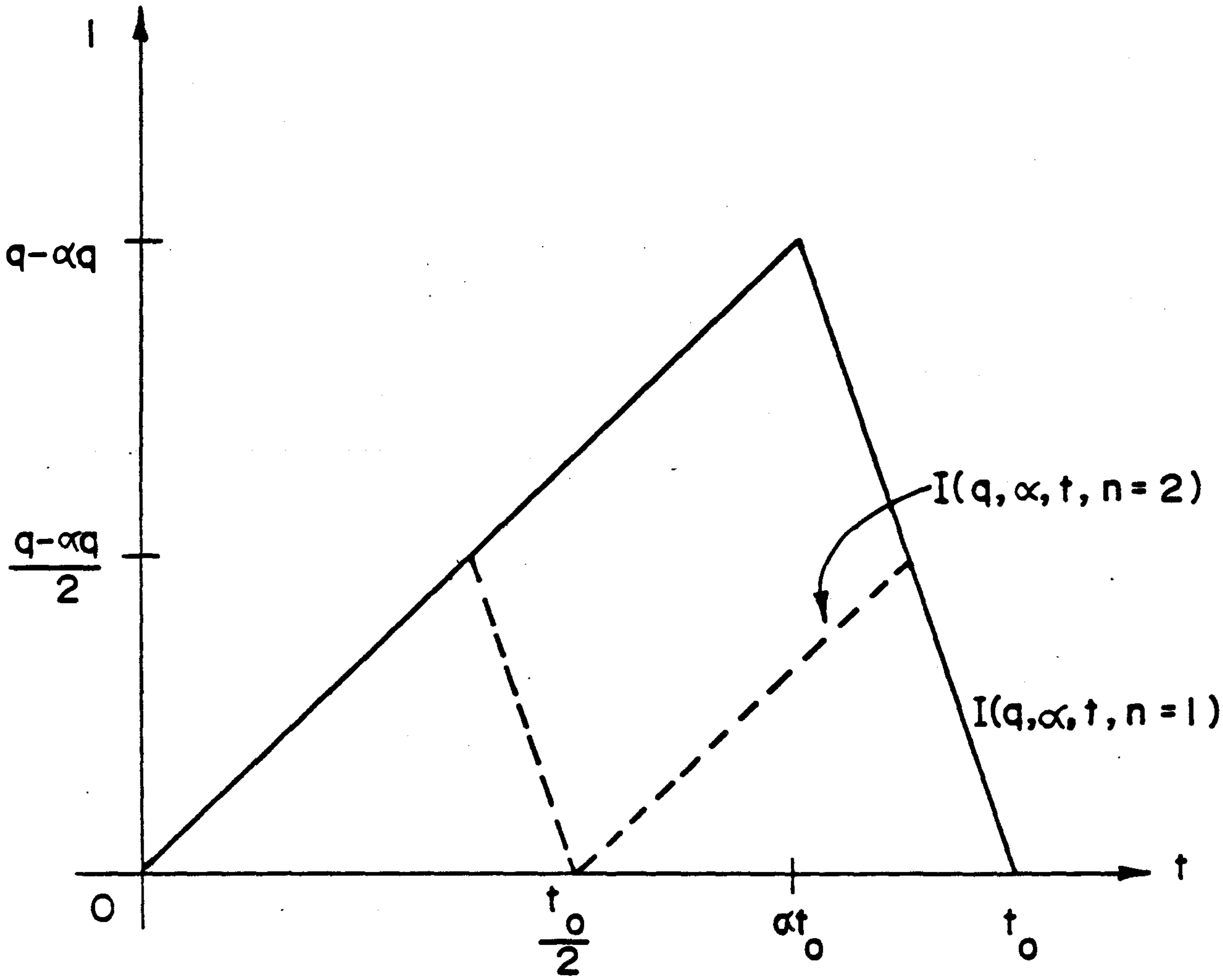


Figure 4