Michael T. Maloney and Robert E. McCormick
Clemson University

A Theory of Cost and Intermittent Production*

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. Transactions costs associated with intermittent production induce the firm to diversify its product line. Moreover, the theory shows that if economies of scale exist it is because of these costs. This implies that economies of scale are an economic not a technological phenomenon.

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. What we do in this paper is detail 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.

* We are indebted to James Buchanan, Ken French, Michael Bradley, Dave Haddock, John Long, Walter Oi, Dan Orr, Robert Tollison, Bruce Yandle, Jerry Zimmerman, workshop participants at the Center for the Study of Public Choice, VPI&SU, and an anonymous referee for helpful comments and suggestions on earlier versions of this paper. The usual caveat applies.

(*Journal of Business*, 1983, vol. 56, no. 2)

0021-9398/83/5602-0003$01.50

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 idling 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] For instance, GM has developed an assembly plant that handles three sizes of car frames. Thus, production of its various sized cars can be adjusted without shutting down a plant.

This theory of intermittent production has additional applications. Consider the monopolist who produces less output than a perfect competitor, other things the same. Does the monopolist produce less at every instant in time by having a smaller capital stock and producing at a slower instantaneous rate of production, or does the monopolist produce at the *same* instantaneous rate as the competitive firm but idle the facility for some portion of time in order to restrict output? Put another way, will a firm that faces a decline in demand reduce its rate of output or simply idle its production facilities intermittently? The model we develop delineates which of these alternative production techniques maximizes the profits of the firm.

The model can also be used to understand the nature and scope of economies of scale. According to our theory, if economies of scale are ever observed, it is not because of lumpiness and indivisibilities, but rather because there are startup and shutdown 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 transactions cost phenomena.

In the next section we define many of the terms employed throughout the paper. Then a zero-transactions-cost model is developed which incorporates time and intermittent production. This construction is 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. A brief conclusion completes the paper.

1. Marketing research has evolved a number of demand-side theories of the multiproduct firm. Our theory provides an additional motivation for multiproduct diversification—cost savings.

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 our presentation and these earlier papers, we first define a number of the terms used in this paper. 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_0 and output level q_0, cost is C_0. Initially we assume that the instantaneous rate of production is constant over the whole time period so that $dq/dt = q_0/t_0$. 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. We 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_0/t_0, 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 in Robinson (1931).[4] For a given time period, at low levels of output and constant production, increasing output reduces marginal and average cost as the firm overcomes lumpiness or indivisibilities in production. Each point on $C_q(\cdot)$ 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_0. (See fig. 1.)

We assume that as the time period increases, both $R_q(\cdot)$ and $C_q(\cdot)$ shift horizontally by the same proportion as the increase in time. For example, $R_q(q, t_0 + \beta t_0)$ and $C_q(q, t_0 + \beta t_0)$ are, respectively, the marginal revenue and marginal cost curves for $(t_0 + \beta t_0)$ time. That is, if the firm can produce q_0 output in, say, 1 month, it can produce $2q_0$ output in 2 months with no change in marginal cost. This assumption implies that $C(q, t)$ is homogeneous of degree 1 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,

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. $C_q(q, t) \equiv \partial C(\cdot)/\partial q$.

4. Robinson (1931) 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.

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

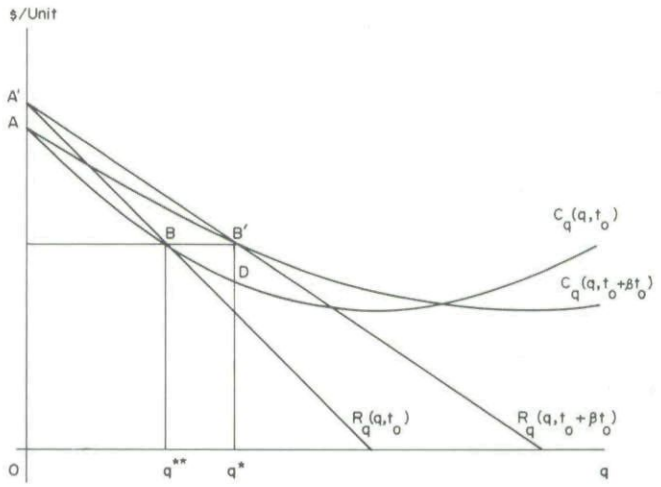

FIG. 1

because we have not changed the total planned volume of production. Whatever the planning horizon of the firm, we hold it constant. The time period t_0 represents a snapshot of the firm, and by increasing the length of time we only expand the exposure period of observation. Hence, there are no real changes within the firm—rate, total planned volume, and costs are unchanged. All we have done is to observe the firm longer, for a time period of length $(1 + \beta)t_0$ rather than the originally considered period t_0.

In the usual presentation, points B and B' in figure 1 satisfy the first- and second-order conditions for profit maximization in t_0 and $(t_0 + \beta t_0)$, respectively. When operating in time period t_0, the firm produces q^{**} and profits are the area $A'BA$. If this firm were to produce q^*, which $= q^{**}(1 + \beta)$, in $(t_0 + \beta t_0)$ time, profit would be $\Pi(t_0 + \beta t_0) = R(q^*, t_0 + \beta t_0) - C(q^*, t_0 + \beta t_0) = A'B'A$. In fact, this solution does not maximize profits. Suppose the firm produced quantity q^* in time t_0 and sold it in time $t_0 + \beta t_0$. 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 1 in output and time,
2. All resources are costlessly liquidated at the end of the production period,
3. The firm can store output costlessly,

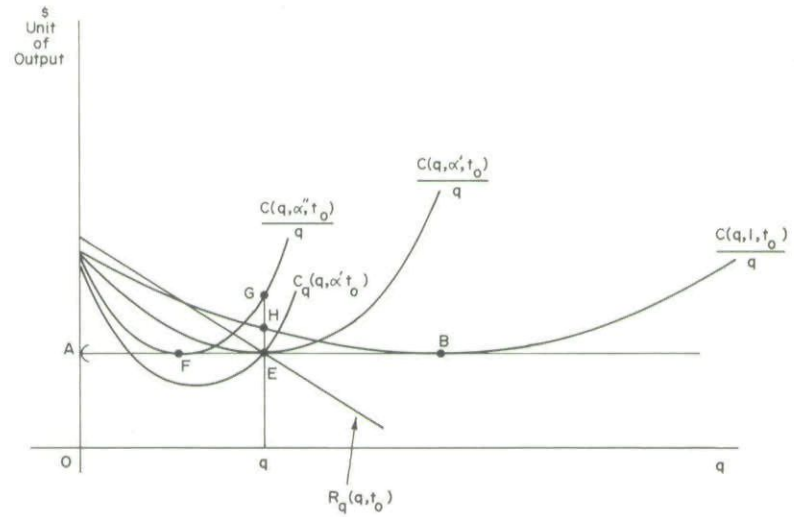

FIG. 2

4. There are no extra costs associated with startup and shutdown, and
5. The discount rate is zero.

As the total cost function is homogeneous of degree 1 in output and time, it is easily shown that $C_q(\cdot)$ and $C(\cdot)/q$ are both homogeneous of degree 0 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 exists for $0 < \alpha \leq 1$. Each α determines an average cost curve for uninterrupted continuous production over the period αt_0, 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_0: $C = C(q, \alpha, t_0)$.

The profit function for the firm is $\Pi(t_0) = R(q, t_0) - C(q, \alpha, t_0)$. Maximizing profit with respect to output and the period of production, subject to the constraint $\alpha \leq 1$, yields

$$R_q(\cdot) = C_q(\cdot) \tag{1}$$

and

$$-C_\alpha(\cdot) = \kappa, \tag{2}$$

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

$$-C_{\alpha\alpha}(\cdot) < 0 \tag{3}$$

and

$$-C_{\alpha q}^2(\cdot) < C_{\alpha\alpha}(\cdot)[R_{qq}(\cdot) - C_{qq}(\cdot)]. \tag{4}$$

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 one, 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, not only that a monopolist operates at the minimum average cost during the period of production, but that 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 because costs and revenues are not paired. Production costs are borne only during the period of production, while revenues are received continuously over t_0. Consequently, an increase in the interest rate induces the firm to increase the period of production. As the firm stands at time 0 and plans production over time t_0, an increase in the discount rate will force the firm to forgo some current production (costs) whose present discounted value is relatively lower. Hicks (1968, pp. 213–17) concludes much the same in his treatment of interest and production. The problem is treated more carefully in the Appendix below.

Storage, Startup, and Shutdown Costs

Storage costs undermine the firm's incentive to employ intermittent production. When these costs are important, the firm reacts by choosing numerous shutdowns of short duration, which drives storage costs toward zero. Startup and shutdown 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 $q/\alpha t_0 - q/t_0$ per instant of time while production proceeds and at the rate of $-q/t_0$ 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.

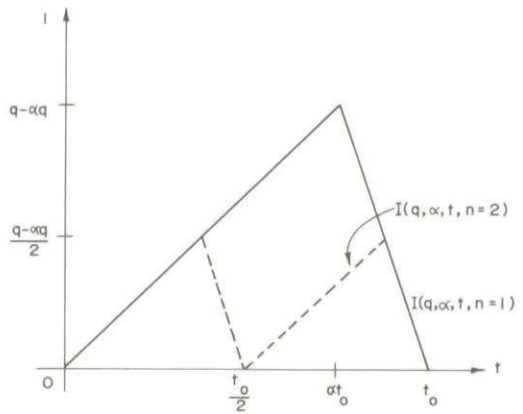

FIG. 3

Inventories are

$$I(t) = (q/\alpha t_0 - q/t_0)t, \qquad t \leq \alpha t_0,$$
$$= q - (q/t_0)t, \qquad t \geq \alpha t_0,$$

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 the first portion of t_0, its storage bill will be $kq(1 - \alpha)t_0/2$. 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 $\alpha t_0/2$ instead of one of αt_0. In fact, the firm can break the process into many production periods, n, each of length $\alpha t_0/n$ accompanied by a shutdown of $(1 - \alpha)t_0/n$.[6] Storage costs are $kq(1 - \alpha)t_0/2n$ and diminish to zero as the number of production breaks goes to infinity. However, starting and stopping production may add costs: label these S_1 and S_2. The optimal number of startups and shutdowns occurs when the reduction in storage costs is equal to startup cost plus shutdown cost, that is, where

$$kq(1 - \alpha)t_0/2n^2 = S_1 + S_2.^7 \tag{5}$$

Because storage costs approach zero as production is divided into many short periods, the average cost curve of the firm remains flat

6. When the firm uses multiple startups its inventories for each of these startup-shutdown cycles are:

$$I(t) = (q/\alpha t_0 - q/t_0)t, \qquad t \leq \alpha t_0/n,$$
$$= q/n - (q/t_0)t, \qquad t \geq \alpha t_0/n,$$

or, in general, $I = (q, \alpha, t, n)$, $I_q > 0$, $I_\alpha > 0$, and $I_n < 0$.

7. Obviously, the choice of n is endogenous with the choice of q and α, and profits are only maximized by simultaneously choosing the three. However, the principle comparative static result is identified by eq. (5), as it is one of the first-order conditions when storage, startup, and shutdown costs are included in the model. Formally, the sign of the derivative $dn/d(S_1 + S_2)$ is negative, as it is the ratio of the second- and third-order principle minors of the hessian of second partials, and these must alternate in sign if the sufficient second-order conditions for a maximum are met. By similar analysis, dn/dk is positive, which says that as the cost of storage goes up the optimal level of inventories falls. If k is an increasing function of I then the preceding comparative static results hold a fortiori. The other comparative static results, $dq/d(S_1 + S_2)$ and $d\alpha/d(S_1 + S_2)$, are indeterminate. This result can be anticipated from a close inspection of fig. 2. If marginal revenue cuts marginal cost for $\alpha = 1$ (MC for continuous production; not shown in fig.) where it lies below the marginal cost for $S_1 + S_2 = 0$, locus $(A, B]$, then q will fall as $S_1 + S_2$ goes to zero. Conversely, if MR cuts $MC(\alpha = 1)$ where it lies above $MC(S_1 + S_2 = 0)$, the opposite holds. The discussion in the next section is a less formal examination of these effects. The term $[-kq(1 - \alpha)t_0/2n(n - 1)]$ is the difference in storage costs for integer reductions in n. In the limit as $\Delta n \to 0$, $\partial S/\partial n = -kq(1 - \alpha)t_0/2n^2$.

even when we incorporate storage so long as startup and shutdown costs are zero. Also, as the firm breaks the time period t_0 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_0, production and sales are nearly paired, and the firm will not be as concerned about discounting the future revenues because they are received almost simultaneously with costs.

Summary

The period of production is a technique that isolates 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 startup/shutdown costs, demand only determines the period of production and the volume of production, not the production rate.

Startup and shutdown 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 1 hour of work than for a 40-hour week, or when it is not free to travel to the

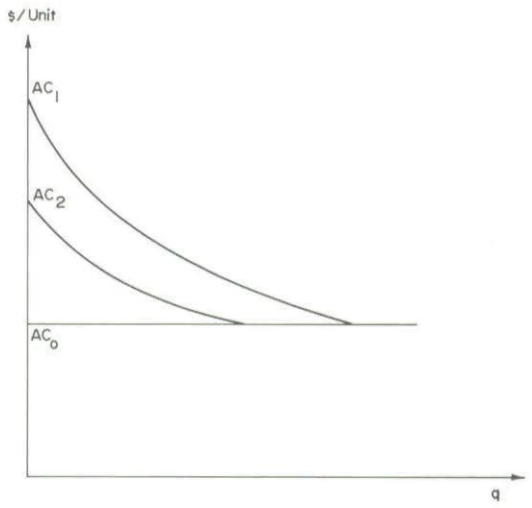

FIG. 4

capital rental store, the average cost curve need no longer be perfectly elastic.[8]

Consider figure 4 where for comparison purposes we have drawn three cost curves. AC_0 is our 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 startup and shutdown costs are so 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_0 and AC_1, embodies costly intermittent production and characterizes the firm.[9] Average cost declines in AC_2 because the firm is unable to enjoy fully returns to scale due to startup and shutdown costs. In the limit, as startup and shutdown costs approach zero and as the number of startups goes to infinity, the cost curve approaches AC_0. As startup costs increase, the optimal number of interruptions diminishes toward zero and consequently the cost curve rises, approaches AC_1, and equals AC_1 when $n = 1$.

8. 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.

9. Cost minimization, with constant output and for $\alpha < 1$, reveals that $d\alpha/d(S_1 + S_2) = (S_{n\alpha}/\Delta) > 0$ because $S_{n\alpha}$ is positive and Δ, the hessian of second partials, is also positive if the sufficient second-order conditions are satisfied.

This means that economies of scale are an economic, not a technological phenomenon. Costly startups and shutdowns 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, startup and shutdown 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 shutdowns 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.[10] The importance of this result is that we expect to see firms adjusting their period of production as well as their output rate when demand fluctuates.

IV. The Multiproduct Firm

Procter 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 because 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.[11]

Procter 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. Then 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

10. 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 shutdowns (unemployment) may be a predictable macroeconomic response to cost-minimizing intermittent production.

11. 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 & Gamble employee.

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 production of another. We expect the merger of Tide with other soaps until the production facility is used full time. This merger, at one plant, makes the transfer of resources associated with intermittent production less expensive.[12]

The characterization of this phenomenon in terms of a single plant results from our focus on capital as the input most costly to startup and shutdown. 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.

There may be extra organizational costs associated with multiproduct firms, in which case we expect the firm to add products until the extra organizational costs equal the marginal cost savings from fewer startups. Intuition suggests that these startup costs will be lowest within the firm when the productive resources have to undergo the slightest physical change.[13] 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 saving from employing intermittent production provides an additional motivation for the firm to undertake product expansion. This is another version of the Coasian firm (see Coase 1937) where transactions within the firm are cheaper than market transac-

12. 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, our discussion 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.

13. Undiversified firms using intermittent production will seek product markets in which to expand. Firms that have systematic prolonged periods of inactivity and firms whose inventories follow the cyclical pattern in fig. 3 are expected to develop new products, and they may merge with other firms that have the same problem. Such mergers will lead to physical consolidation of production, shorter idle periods, and higher profits, but a cyclical pattern in inventories will persist. These mergers will lead to lower costs and output prices, higher stock prices of both firms, and lower stock prices of rivals. Physical diversification will also lower the beta of the postmerger stock return because the firm has a wider range of production opportunities. Mandelker (1974) finds, for mergers of all types between 1941 and 1962, that postmerger betas are approximately 8% lower than premerger betas but not in a statistically reliable sense. We expect that a subsample of his firms which have cyclical production before merger will have a larger premerger-postmerger difference in beta.

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.

tions. Our model 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 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 resources that can only be used efficiently in an intensive manner.

Thus far, our 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 we consider randomly fluctuating demand and U-shaped average variable cost curves, 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. Our 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, we expect 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 wheat falls, farmers transfer production to cattle, hay, or soybeans. They do not choose a slower-growing variety of wheat.

V. Conclusion

We have developed 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, startup and shutdown costs reduce the cost savings from this strategy. This means that economies of scale are a transactions cost, not a technological phenomenon.

In some cases firms find it profitable simply to idle their facilities from time to time while selling off accumulated stocks. In other cases,

startup and shutdown costs are great enough 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 startup and shutdown 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.

Appendix

If we introduce continuous discounting of costs and revenues because the two are not paired, then the present value to the firm of the production process over time t_0 is given by

$$W = \int_0^{t_0} [R(q, t_0)/t_0]\, e^{-ru}\, du - \int_0^{\alpha t_0} [C(q, \alpha, t_0)/\alpha t_0]\, e^{-ru}\, du, \tag{A1}$$

because the instantaneous revenues and costs are $R(q, t_0)$ and $C(q, \alpha, t_0)$, respectively. Integration of (A1) yields:

$$W = [R(q, t_0)/t_0] \cdot [(1 - e^{-rt_0})/r] - [C(q, \alpha, t_0)/\alpha t_0] \cdot [(1 - e^{-r\alpha t_0})/r]. \tag{A2}$$

The first-order conditions necessary to maximize (A2) with respect to q and α are given by

$$[(1 - e^{-rt_0})/rt_0]\, R_q(\cdot) = [(1 - e^{-r\alpha t_0})/r\alpha t_0]\, C_q(\cdot) \tag{A3}$$

and

$$(1/\alpha)[(1/\alpha t_0)C(\cdot)(1 - e^{-r\alpha t_0})(1/r) - (1/t_0)C_\alpha(\cdot)(1 - e^{-r\alpha t_0})(1/r) - C(\cdot)e^{-r\alpha t_0}] = 0. \tag{A4}$$

Equation (A4) is our new first-order-condition corollary to $-C_\alpha(\cdot) = 0$. From (A4) it is easily shown that $C_\alpha(\cdot)$ is now positive. Thus, the optimal period of production is greater with discounting (positive interest rates) than without. For simplicity we have placed all production at the beginning (first αt_0) of the production period. As noted in Section III, this is not necessarily cost minimizing. To the extent that there is more than one shutdown in the time period, the discounting role becomes less important as costs and revenues become more nearly paired.

References

Alchian, A. A. 1959. Costs and outputs. In M. Abramovitz et al., *The Allocation of Economic Resources*. Stanford, Calif.: Stanford University Press. (Reprinted in *Readings in Microeconomics*, 2d ed., ed. W. Breit and H. M. Hochman [Hinsdale, Ill.: Dryden, 1971].)

Alchian, A. A., and Demsetz, H. 1972. Production, information costs, and economic organization. *American Economic Review* 62 (December): 777–95.

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

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

Coase, R. 1937. The nature of the firm. *Economica* 4:386–405.

Gort, M. 1962. *Diversification and Intergration in American Industry*. Princeton, N.J.: Princeton University Press, for the National Bureau of Economic Research.

Hicks, J. R. 1968. *Value and Capital*. 2d ed. Oxford: Oxford University Press.

Hirshleifer, J. 1962. The firm's cost function: a successful reconstruction. *Journal of Business* 35 (July): 235–55.

Mandelker, G. 1974. Risk and return: the case of merging firms. *Journal of Financial Economics* 1:303–35.

Orr, D. 1964. Costs and outputs: an appraisal of dynamic aspects. *Journal of Business* 37 (January): 51–60.

Pfouts, R. W. 1961. The theory of cost and production in the multi-product firm. *Econometrica* 29 (October): 650–58.

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

Stigler, G. J. 1951. The division of labor is limited by the extent of the market. *Journal of Political Economy* 59 (June): 185–93.