Communication and Inventory as Substitutes in Organizing Production*

Paul Milgrom and *John Roberts*
Stanford University, Stanford, CA, USA

Abstract

A major organizational design decision for manufacturing firms is the extent to which production should be to stock versus to order: inventories and communication with customers are substitutes. We show that profits are convex in the share of the market supplied from inventory. Thus firms will tend to specialize in one mode of organization. We examine how this choice depends on market size, the level and variability of demand, the costs of communication, price levels, production costs, and the costs of expanding product lines. The results are consistent with observed patterns in several industries.

I. Introduction

The publishing industry in England in the eighteenth and early nineteenth century commonly operated on a produce-to-order basis, employing the "subscription" system; see e.g., Blond (1971, p. 52). Before a book was published, the publisher (who was often a bookseller) would seek orders for it from potential readers. If the demand then justified publishing the book, the initial printing run would be geared to meeting these advance orders. In contrast, while advance orders may still be sought,[1] it has now become standard to commit to publication and to determine the size of the initial printing in anticipation of the actual realization of demand. Orders, as they materialize, are then filled from inventory.

Something of the reverse shift has recently been observed in connection with "flexible manufacturing", under which the same workers, machines and facilities are able to shift quickly between producing different products. Often associated with adoption of these methods is a shift from

[1] This is especially true for certain very expensive reference books.

*We are grateful to Joseph von R. Farrell, J. Michael Harrison, Michael Riordan and two anonymous referees for their suggestions, to the National Science Foundation and the John Simon Guggenheim Memorial Foundation for their financial support, and to the Economics Department at U.C. Berkeley for its hospitality.

producing to inventory to producing more to order, especially at the level of components or intermediate goods. The celebrated *kanban* system, in which information on demand is passed up from one stage of production to earlier ones, and "just-in-time" methods, with inventories of work in progress being minimized by having needed parts made available just as they are needed, are manifestations of this shift; on these matters, see Aggarwal (1985).

These examples suggest that inventories and communication of information about demand are substitutes for one another in production. They also suggest some tendency to organize so as to employ only one or the other — firms either make to stock or make to order. However, shifts from one alternative to the other do occur.

The purpose of this paper is to explore the nature of the substitution possibilities between communication and inventories and to study how the optimal use of these substitutes — and, correspondingly, the organization of the firm — depends on technology, cost and demand conditions.

The make-to-stock versus make-to-order question is a fundamental one in organizing production, but one that has been subject to relatively little formal modeling in either the economics or operations management literatures.[2] Moreover, this particular issue is representative of a much broader question in organizational design.

Galbraith (1973) has argued that the degree of uncertainty involved in task execution is the key determining factor in designing effective organizations. He also identified a number of broad strategies for dealing with uncertainty. Using inventories and learning demand before undertaking production are examples of two of these. The first involves the provision of "slack", that is, the conscious lowering of output objectives or the allocation of resources above what would be needed in expectation (or in some "base-case" scenario) that permit tasks to be carried out in the event of unusually high demands on the organization. The second involves increasing the vertical flow of information in the organization so as to facilitate forecasting and planning that permit a better match between the resources allocated to a task and those needed to complete it.

These alternatives have different costs. Using slack means that some of the time costly inputs will be wasted *ex post*, and unless enough resources are made available to provide for every contingency, there will still be occasions when the tasks are not successfully executed. Increasing the information flow, on the other hand, involves costs in obtaining, transmitting and processing the information and in formulating, communicating and implementing the plans based on it. Thus, the relative efficiencies of

[2] One interesting exception is Lim (1981). Dudley and Lasserre (1986) provides empirical evidence that information substitutes for inventories.

these alternatives will depend on the circumstances facing the organization, and the issue becomes one of determining how the organization will optimally mix its use of the two as these circumstances change.

We illustrate the tradeoffs between these differing approaches to coping with uncertainty in a model of production planning. A firm producing a variety of products faces random demands for each of them. We model these demands as being independently and normally distributed with known mean and variance. This could arise if demand for each good was the sum of the independent random demands of many small customers.[3] The firm has the option of learning the actual, realized demand from any fraction of the customers for each product, with the cost of doing so being linear in the fraction of the market surveyed. Demand from any portion of a market not surveyed is then served by producing to inventory.

Assuming that orders from surveyed customers will be met, the firm must decide what fraction of each of its markets to survey and how much to produce for inventory of each product in order to maximize expected profits.

Our central result (Theorem 2) is that this profit maximization problem involves an essential nonconcavity: holding the probability of stock-outs (the "service level") fixed, profits are a convex function of the fraction of each market to be surveyed. Thus, either all production is for inventory or all is to order. This nonconcavity arises because the possibility of pooling the uncertain demands of different individuals induces economies of scale in inventory as the size of the market which the inventories serve is increased.

We also model the firm's choice of the set of products it will offer. This choice is made knowing how the distribution of demands depends on the actual product line chosen but before it is possible to gather information on realized quantities demanded. Further, we assume that the net price received by the firm is an increasing function the size of the product line. The choice of the optimal product line then involves trading off the additional design and set-up costs of adding to the line against the resultant improved price. Moreover, a broader product line means that the market for each product actually produced is narrower, and this is costly when producing for inventory because of the lost pooling. In turn, this has implications for the relative profitability of producing to realized demand

[3] In this context, knowledge of the mean and variance could have come from past experience or from market research which identified these parameters without identifying the demands of individual customers. A topic for future work involves the issues arising when the parameters of the market demand distribution are themselves unknown and must be learned, but in the present model the only demand uncertainties are those arising from the independent decisions of the individual customers.

(the "communications regime") or in anticipation of demand (the "inventory regime").

In this context we explore how the choice between making to order or to stock depends on the various parameters of the model, including: the mean level and variability of per customer demand; the overall number of customers; the price received; and the costs of production, of expanding the product line and of surveying customers. We do this comparative statics analysis for both the cases where the number of products offered is exogenously specified (Theorem 3) and where it is optimally adjusted to the regime chosen (Theorem 5). We also obtain comparative statics results on the choice of the size of the product line within each regime (Theorem 4).

This modeling seems to capture salient aspects of a number of actual production processes. First would be the book publishing example mentioned above. A second comes from the semiconductor industry.[4] There is potential demand for an immense variety of semi-conductor chips, each of which carries out a particular logic, and the quantity demanded of any of these is random from the point of view of the producers of these chips. Producers respond to this uncertainty by creating a limited variety of "gate arrays", which are, in essence, partially completed chips which can be customized at some cost to perform any one of a variety of specific tasks as actual demands arise. In this context, the standard practice is to produce gate arrays to inventory. However, industry executives have expressed great interest in obtaining information from customers about their demands so as to economize on these inventories.

Scheduling courses in a university department presents another possible example. Student interests in courses vary widely and in a partially unpredictable fashion. The problem is to determine which courses to design and offer, and then how many spaces to make available in each. One approach to the scheduling problem is simply to decide on these matters in advance; the other is to attempt to gather demand information before determining the number of spaces to assign to each course. The former results in over- and under-subscribed sections; the latter involves costs in gathering reliable information about students' intentions.

A final example comes from the restaurant business, where the firm must decide on its menu and on whether to produce each dish to order or to hold food already prepared to be served as demand arises.

Our comparative statics results are consistent with and illuminate a variety of observed behavior in each of these examples. The formal model is developed in the following section, as are the results. The final section

[4] We are indebted to J. Michael Harrison for conversations on this topic.

contains some further discussion on the examples in light of the results and also reports our conclusions.

II. The Model and Results

The model is one in which there are many potential products aimed at a total customer market of "size" m. Each customer's purchases are a small fraction of the total sales of the firm. Given any specified list of products offered by the firm, each customer plans to buy a single variety. The individual customers' demands[5] are asssumed to be statistically independent and identically distributed, with finite variance. Therefore, over any fraction λ of the market, demand is approximately normally distributed with mean and variance each proportional to $m\lambda$, the size of the market segment. We take this approximation to be exact in the following analysis, even though this injects the possibility that demand for some offered varieties could be negative.

The firm must first select the number, K, and the design of the product varieties it will offer. Each additional product offered involves incurring a cost $c>0$ for product design, machine set-ups, etc., but also yields a higher average price received for each product. Thus, as long as the price that might be received remains bounded, only a finite variety of products will be offered. Once the number and design of products to be offered has been determined, the set of customers is partitioned into K submarkets, and, given our previous assumptions, the demands in these submarkets are statistically independent.

The chief economic content of this independence condition is that if demand for good k exceeds the amount available, this excess demand cannot be transferred to another good. This actually appears to be the case in the gate-array example mentioned earlier, and would hold generally when the products in question are intermediate goods or components and the final goods in which they are used can be designed to be compatible with at most one of these intermediate products. On the other hand, in the course section planning and restaurant menu design problems, some shifting of demand is possible. In such cases, the particular model used here fits less well, but aspects of the basic structure and results remain, as does the essential feature that stocking-out of a customer's most preferred product is costly to the firm.

We assume that the price received per unit sold, $p(K)$, is an increasing, concave, bounded function of K, the number of products offered, and that

---

[5] To simplify matters, the dependence of quantity demanded on price is ignored. This may reflect the price's being competitively determined or set through long-term contracts.

the firm recognizes this effect in its decision-making. We also assume that, for any K, the set of customers interested in each variety is of size m/K.

Production of a unit of output of any good costs γ, given that the design and tooling costs c for this good have already been met.

Finally, the firm has the option of surveying any fraction of its customers and thereby learning perfectly what their demands will be. We assume that these survey costs are of the form $S+\lambda s$, where λ is the size of the set of customers surveyed, $S \geq 0$ are fixed costs which are incurred only if some surveying is actually done, and $s \geq 0$ is the marginal cost of surveying an additional fraction of the market. The constant marginal cost assumption seems appropriate in a survey context when all customers are *ex ante* identical.

With these assumptions, the problem of the firm is to maximize expected profits with respect to the number K of products to offer, the fraction α of each product market to survey, and the output level x to produce to meet demand from the unsurveyed customers in each segment.

Before expressing this problem formally, a change of variables is convenient. Given α and K, define

$$z = \frac{x-\mu(1-\alpha)m/K}{\sigma\sqrt{(1-\alpha)m/K}};$$

z is now a standard normal variate interpretable as the number of standard deviations by which production for inventory in any submarket differs from the expected demand of the unsurveyed customers. It satisfies $F(x)=\Phi(z)$, where F is the distribution of the original demand (assumed, for each value of α, m and K, to be normal with mean $(1-\alpha)\mu m/K$ and variance $(1-\alpha)\sigma^2 m/K$ in each submarket) and Φ is the standard normal CDF with density ϕ. The firm's problem can now be written as

$$\begin{aligned}\max_{K,\alpha,Z}\; & K[p(K)-\gamma]\int_{-\infty}^{\infty}\left[\frac{\alpha\mu m}{K}+\sigma\sqrt{\frac{\alpha m}{K}}\,y\right]\phi(y)\,\mathrm{d}y \\ & + Kp(K)\int_{-\infty}^{\infty}\left[\frac{(1-\alpha)\mu m}{K}+\sigma\sqrt{\frac{(1-\alpha)m}{K}}\min(z,y)\right]\phi(y)\,\mathrm{d}y \\ & -\gamma\left[\frac{(1-\alpha)\mu m}{K}+\sigma\sqrt{\frac{(1-\alpha)m}{K}}\,z\right]K-\alpha sm-\delta_\alpha S-cK,\end{aligned}$$

where $\delta_\alpha=0$ if $\alpha=0$ and $\delta_\alpha=1$ if $\alpha>0$.

The first term of this expression is the expected contribution to profit and overhead from the surveyed demand; it is K times the markup over direct production cost times the expected demand on each of the surveyed subsets (each of which is of size $\alpha m/K$).[6] The "min" in the next term reflects the fact that sales to unsurveyed customers in each submarket are the smaller of realized demand and the amount of output produced for inventory in anticipation of this demand. The remaining terms are, respectively, the production cost of goods produced for inventory, the marginal and fixed costs of gathering information, and the design and set-up costs for the products.

Theorem 1. *The optimal level of production for inventory, expressed in terms of standard deviations from mean demand, is a function of K and γ, but is independent of m, μ, σ^2, α, S and s. It is given by*

$$z=\beta(K,\gamma)=\Phi^{-1}(1-\gamma/p(K)).$$

The optimal level of output in unit terms is thus

$$x=\frac{(1-\alpha)\mu m}{K}+\beta(K,\gamma)\sigma\sqrt{\frac{(1-\alpha)m}{K}}.$$

Proof: Given K and α, the first-order condition,

$$p(K)[1-\Phi(z)]=\gamma,$$

equates the expected marginal revenue[7] of producing an extra unit to the marginal cost. The second derivative of the objective function with respect to z is

$$-\sigma\sqrt{\frac{(1-\alpha)m}{K}}\,\phi(z)<0,$$

so the optimum is characterized by the first-order condition. Q.E.D.

Simplifying the expression for profits by integrating yields the following for the firm's problem:

$$\max_{K,\alpha,Z} K(p(K)-\gamma)\frac{\alpha\mu m}{K}$$
$$+Kp(K)\left[\frac{(1-\alpha)\mu m}{K}+\sigma\sqrt{\frac{(1-\alpha)m}{K}}\int_{-\infty}^{\infty}\min(z,y)\phi(y)\,\mathrm{d}y\right]$$
$$-\gamma(1-\alpha)\mu m-\gamma\sigma\sqrt{(1-\alpha)mK}z-\alpha sm-\delta_\alpha S-cK.$$

[6] The form of this expression assumes that surveyed demand is met. This is optimal, given $p(K)>\gamma$.

[7] The expected marginal revenue is less than price because extra production is sold only if demand exceeds production, which occurs with probability $1-\Phi(z)$.

Note that

$$\int_{-\infty}^{\infty} \min(z, y)\phi(y)\,\mathrm{d}y < \int_{-\infty}^{\infty} y\phi(y)\,\mathrm{d}y = 0.$$

Thus, the firm's profit is the sum of terms that are linear in α and terms that are strictly convex in α, and so is itself strictly convex in the fraction of the market surveyed. Consequently, we have our fundamental result:

Theorem 2. *The optimal choice of α is always either $\alpha = 0$ or $\alpha = 1$.*

The intuition here is as simple as the proof. When producing to inventories, there are generally economies of scale in the weak sense that profits increase more than proportionally with market size (subadditivity); with the normally distributed random demand considered here, there are economies of scale in the stronger sense that profits are a strictly convex function of market size. Meanwhile, when producing to order, profits are simply proportional to market size. Thus, if it pays to serve any portion of the market via production to inventory, it is optimal to serve the entire market in this fashion. Thus, too, although communication and inventories are substitutes in dealing with uncertainty, only one or the other will optimally be used.[8]

The remainder of our analysis is concerned with exploring the conditions determining which of the two alternative regimes is selected. It will simplify matters to consider first the effects of changing various parameters, given K, then to incorporate the adjustments in K.

Note that expected production and sales under the communications regime ($\alpha = 1$) are μm, so for any K the expected profits are

$$\Pi^{C}(K) = (p(K) - \gamma)\mu m - sm - S - cK.$$

Under the inventory regime ($\alpha = 0$), if the production level of each product is x, then expected sales volume per market segment is

$$\int_{-\infty}^{x} yf(y)\,\mathrm{d}y + (1 - F(x))x,$$

where F is the CDF for a normal random variable with mean $\mu m/K$ and variance $\sigma^2 m/K$ and f is the corresponding density function. Thus, expected profits are

[8] Note that this result is sensitive to the *ex ante* homogeneity of the customers. If customers vary in size or in the costs of learning their demands, it is optimal to survey large customers and to serve the others from inventories.

$$Kp(K)\left[\int_{-\infty}^{x} yf(y)\,\mathrm{d}y+(1-F(x))x\right]-K\gamma x-cK.$$

By Theorem 1, at the optimum production, $F(x^*)=\Phi(z^*)=1-\gamma/p(K)$, so maximized expected profits under the inventory regime are

$$\Pi^I(K)=Kp(K)\int_{-\infty}^{x^*} yf(y)\,\mathrm{d}y-cK$$

$$=Kp(K)\int_{-\infty}^{z^*}\left[\frac{\mu m}{K}+\sigma\sqrt{\frac{m}{K}}\,t\right]\phi(t)\,\mathrm{d}t-cK.$$

Because $t\phi(t)=-\phi'(t)$, we obtain

$$\Pi^I(K)=Kp(K)\left[\frac{\mu m}{K}\Phi(z^*)-\sigma\sqrt{\frac{m}{K}}\,\phi(z^*)\right]-cK$$

$$=\mu m(p(K)-\gamma)-p(K)\sigma\sqrt{mK}\,\phi(z^*)-cK.$$

The corresponding expected sales volume in terms of z^* is

$$\mu m-\sigma\sqrt{mK}\,\phi(z^*)+\gamma z^*\sigma\sqrt{mK}/p(K).$$

These expressions are the basis for obtaining the comparative statics results with a fixed value of K.

First, suppose $p(K)$ increases, and consider the value of profits under the communication regime and under the inventory regime after optimizing over the output level. In each case, the derivative of optimized profits with respect to $p(K)$ is the expected sales volume. (In the inventories case this result uses the envelope theorem.) But expected sales under the communication regime are μm, while under the inventory regime they are $K\int\min(x,y)f(y)\,\mathrm{d}y<K\int yf(y)\,\mathrm{d}y=\mu m$. Thus, if collecting information was optimal for a given value of $p(K)$, it will also be optimal for higher prices.

Second, suppose γ, the production cost, increases. The derivative of the (optimized) profits with respect to γ under either regime is the negative of the expected production level. Using inventories, this is $-F^{-1}(1-\gamma/p(K))$, while under the communication regime it is $-\mu m$. Thus, which regime is favored by an increase in γ depends on which one involves the larger level of production, i.e., on whether $1-\gamma/p(K)$ is greater or less than $\frac{1}{2}$. If $\gamma<p(K)/2$, then $F^{-1}(1-\gamma/p(K))>\mu m/K$, so the communication regime is favored. On the other hand, if costs are relatively high, and so output in the inventory regime is less than the expected demand, a further increase in costs favors producing for inventory over gathering information.

A simultaneous increase in γ and $p(K)$, holding their ratio constant, favors the communications regime. This can be verified by direct computation. The sense of this result is that it may be worthwhile to obtain information on the demand for high-value items and yet not be optimal to set up complicated communications systems to learn the demand for low-value ones, even when the percentage markup is the same: one might produce supercomputers to order, but nuts and bolts for inventory.

An increase in m, the "size" of the market, should be interpreted as a growth in the number of customers, each of whose demand remains unchanged. Such market growth tends to favor the inventory regime. The derivative of optimized profits under the communications regime with respect to m is $\mu(p(K)-\gamma)-s$, while that under the inventory regime is

$$\mu(p(K)-\gamma)-(p(K)\phi(\beta(p(K),\gamma))\sqrt{K})/(2\sqrt{m}),$$

where $x=(\mu m/K)+\beta(p(K),\gamma)\sigma\sqrt{m/K}$ is the optimal output for each of the K market segments. Because $\beta(p(K),\gamma)$ is independent of m, the derivative of optimized profits under the inventory regime is of the form $\mu(p(K)-\gamma)-H(K,\sigma)/\sqrt{m}$. Consequently, this derivative exceeds that of profits under the communications regime for large m, so further increases in m favor using inventories rather than gathering demand information. Moreover, since asymptotically the derivative of profits when producing for inventory exceeds that when producing to order by an amount s, the marginal survey cost, in large enough markets firms will produce to inventories instead of surveying customers.[9]

Changing μ has no effect on the inventory/communication choice, given the size of the product line, because effectively this is a deterministic shift in demand. Increasing σ^2 favors the communication regime because this leaves expected profits unchanged under this regime but lowers them when inventories are used: $\partial\Pi^I/\partial\sigma=-p(K)\sqrt{mK}\phi(\beta)<0$. Increases in the fixed or variable costs of surveying clearly favor the inventory regime. Finally, exogenous increases in the given, common level of K favor the communication regime. To see this, note first that $\partial\Pi^C/\partial K=p'(K)\mu m-c$, i.e., $p'(K)$ times expected sales volume less the direct cost of expanding the product line. In the inventory regime,

$$\partial\Pi^I/\partial K=p'(K)\cdot(\text{Expected sales volume})-\gamma p'\beta\sigma\sqrt{mK}/p(K)$$
$$-p\sigma\phi(\beta)\sqrt{m}/2\sqrt{K}-p\sigma\sqrt{mK}\phi'(\beta)(\partial\beta/\partial p)p'-c.$$

[9] Profits are zero at $m=0$ in both regimes, but are negative when using inventories for small values of m. This latter point is an artifact of our ignoring nonnegativity constraints: when the number of customers is small, the normal distribution is a poor model for demand.

Using $\phi' = -\beta\phi$ and $\partial\beta/\partial p = \gamma/p^2\phi$, this becomes

$$\partial\Pi^I/\partial K = p'(K)(\text{Expected sales volume}) - p\sigma\phi(\beta)\sqrt{m}/(2\sqrt{K}) - c.$$

Given that expected sales volume is always higher in the communications regime, $\partial\Pi^I/\partial K < \partial\Pi^C/\partial K$ for all K. (The extra term in $\partial\Pi^I/\partial K$ reflects the lost economies of scale in inventories that result from increasing the size of the product line.)

Summarizing the above, we have:

Theorem 3. *For any given K, the relative profitability of the communication regime is increased by: (1) increases in the price received; (2) decreases in production costs when $\gamma > p(K)/2$, and increases when $\gamma < p(K)/2$; (3) proportional increases in γ and $p(K)$; (4) increases in the variability of demand, σ; (5) decreases in m, the thickness of the market, when m is not too small; (6) decreases in the cost of surveying, s and S. Changes in μ, the mean demand per customer, leave the relative profitability of the two regimes unchanged. Exogenous increases in the common level of K favor the communication regime.*

These results for given K can be considered as short-run responses if we think of the determination of the product line as a decision that is fixed for an extended period. Moreover, in many situations, K is naturally integer valued, and so in fact it will typically remain constant for small parameter changes. Finally these results help in examining the effects of changing the various parameters when K is adjusted optimally.

Theorem 4. *Assume that output in the inventory regime is optimally adapted to K. Let K^I be the optimal level of K under the inventory regime and let K^C be the optimal level under the communication regime. Then, when they are unique, $K^C \geq K^I$. Thus, when a shift to the communication regime occurs, it will tend to be accompanied by an increase in the size of the product line. Further, K^C is increasing in μ and m, constant with respect to σ^2 and γ, and decreasing in c, while K^I is increasing in μ and decreasing in c. Both are independent of s and S.*

Proof: These results follow by routine calculation from the formulae given above.[10]

Q.E.D.

The dependence of K^I on m, σ^2 and γ is unclear. Treat K as continuous and assume we can use the implicit function theorem. In each case, the sign of $\partial K^I/\partial t$, $(t = m, \sigma^2, \gamma)$, is equal to that of $\partial^2\Pi^I/\partial K\partial t$. The relevant expressions are

---

[10] Note that even if K is a discrete variable, we may, for example, still use the fact that (formally) $d\Pi^C/dK > d\Pi^I/dK$ to infer that $K^C \geq K^I$.

$$\frac{\partial^2 \Pi'}{\partial K\,\partial m} = \mu p' + \frac{p'\sigma\phi}{2}\sqrt{\frac{K}{m}} - \frac{p'\sigma\gamma\beta}{p}\sqrt{\frac{K}{m}} - \frac{p\sigma\phi}{4\sqrt{Km}}$$

$$\frac{\partial^2 \Pi'}{\partial K\,\partial \sigma} = -p'\sqrt{mK}\phi - \frac{p\phi}{2}\sqrt{\frac{m}{K}} + \frac{p'\gamma\beta\sqrt{mK}}{p}$$

and

$$\frac{\partial^2 \Pi'}{\partial K\,\partial \gamma} = -\frac{p'\sigma\gamma\sqrt{mK}}{p^2\phi} - \frac{\sigma\beta\sqrt{m}}{2\sqrt{K}}.$$

We cannot sign these expressions. However, it is clear that if m becomes sufficiently large, K must eventually be increased for optimality, because, given K, as $m \to \infty$, $\partial^2\Pi'/\partial K\,\partial m \to \mu p' > 0$. The logic here is that as the size of the market becomes arbitrarily large with a given product line, the standard deviation of demand becomes negligible relative to the mean. Thus the negative effect of increasing K, which comes through the stochastic difference between production and demand, correspondingly becomes negligible in comparison to the positive effect of raising the price. Also, as γ/p tends to zero, $\partial^2\Pi'/\partial K\,\partial\sigma \to \infty$, while when γ exceeds $p(K)/2$, β is negative and so $\partial^2\Pi'/\partial K\,\partial\sigma < 0$. Thus, for extreme values of the production cost, the dependence of K' on the variability of demand is also determinate. Finally, since $\partial^2\Pi'/\partial K\,\partial\gamma$ is $(-p'\sigma\sqrt{mK}(1-\Phi)/\phi p) - (\sigma\beta\sqrt{m}/2\sqrt{K})$, the sign of $\partial K'/\partial\gamma$ is asymptotically governed by the behavior of $(1-\Phi)/\beta\phi$. For γ/p close to unity (i.e., $\beta \to -\infty$), this expression goes to minus infinity. For γ/p approaching zero, l'Hôpital's rule gives that $(1-\Phi)/\beta\phi \to 0$. Thus, for extreme levels of cost relative to price, $\partial K'/\partial\gamma$ is negative.

Thus, broader product lines are associated with high levels of expected demand per customer and with low costs of expanding the product line. Also, if it is optimal to produce to order, then thicker markets contribute to broadened product lines. This same effect is at work when producing to inventory, provided the market is sufficiently large. Also, if the inventory regime is used, then the breadth of the product line increases with the variability of individual demand and with decreases in the marginal costs of production when these costs are low, while the reverse is the case when γ is high.

We can now examine the full effects of shifts in the various parameters on the relative profitability of the two modes of organization.

Theorem 5: *The relative profitability of the communications regime is increased by (1) uniform increases in the prices, $p(K)$; (2) increases in the marginal cost of output, γ, if and only if $\gamma < p(K')/2$; (3) proportional* *increases in θ, $p(\cdot)$ and c; (4) increases in the variability of per customer demand, σ; (5) decreases in m, the thickness of the market, when m is sufficiently large; (6) decreases in the costs of surveying, s and S; (7) increases in mean per customer demand, μ; (8) decreases in the cost of expanding the product line, c.*

Proof: If K is a continuous variable, then the results follow in a standard fashion, using the envelope theorem. If K is discrete, there will frequently be multiple optima for K in each regime. Note, however, that for any vector of the exogenous parameter values and any direction of change in these, there exist optimal levels of K under each regime that remain constant under small parameter changes in the given direction. This is because there are only a finite number of optimal K values and we can simply take the one giving the greatest derivative of profits with respect to the parameter change. Thus, even with K discrete, we can treat K as (locally) fixed in each regime, then take derivatives of the expressions for profits in each regime in the usual fashion, using the envelope theorem, to obtain the results. Since the optimal level of K^C exceeds that of K^I, the price received in producing to order exceeds that when producing for inventory. As a result, the signs of the derivatives of $\Pi^C - \Pi^I$ with respect to the various parameters are unambiguous except that with respect to m. In this case, an argument like that in Theorem 3 applies. Q.E.D.

Note that the effect of optimizing over K is that the chosen level of K in the inventory regime is smaller than in the communications regime, and this accounts for the differences between Theorems 2 and 4.

III. Discussion, Summary and Conclusions

We opened this essay with two examples. The first involved a shift from making to order to making to stock in book publishing. A number of our results may explain this. First, if the market for these books grew sufficiently large (with increased population and real incomes, or with changing tastes), then the results of Theorem 3 (with K fixed at one — each book is treated as unique) would explain this shift. Moreover, even if we allow the number of different titles published to expand with m (as in Theorem 5), we would still expect the observed shift to be favored.

A second change that plausibly occurred and would have favored this shift would be a reduction in the marginal costs of printing copies of each title through improved printing technology, the mechanization of bindery operations, or reduced paper costs when accompanied by a not-more-than-proportional reduction in price. Theorems 3 and 5 indicate that such a change will favor a shift to the inventory regime. This would seem to be the case under current demand and technological conditions, and probably was in early periods as well (given the high costs of manual typesetting).

The second example concerns the shift to producing to order and use of the Kanban system in connection with flexible manufacturing. Again, our model suggests a number of plausible changes which would have this effect. Improved forecasting techniques and telecommunications technology would lower the costs of gathering information on demand and directly favor a shift to the communication regime. As well, technological or organizational developments that lower the costs of expanding the product line would tend to favor this shift. In particular, reductions in setup costs achieved by flexible work rules and broadly trained employees (so that production workers do the setups rather than standing idle while specialists do this task) would lead to this shift, as would the development of machines that achieve the low production costs previously associated only with special purpose machines but that can, at low cost, be reprogrammed for other tasks. An interesting empirical question is whether these factors are present in the situations where this shift has occurred.

Within the workplace, our model explains the use of production to order for high value components (high p and γ) but to stock for low value components.

We also cited the example of the restaurant business. High prices favor producing to order in our model. Thus, one would expect that expensive restaurants would tend to produce to order from scratch, while less expensive ones would hold inventories of already prepared dishes. One would also expect that this tendency would be present even when preparing and holding a dish for relatively brief periods (within a single day) in anticipation of demand did not contribute to deterioration in quality.

Our model then seems to have some value in explaining aspects of the choice of organizing to produce for inventory versus organizing to produce to order. This is one of the crucial decisions in organizational design.

In terms of the recent trends in economic research on organization issues, it is striking that incentive questions play no role in our analysis; our model is fully a technology-driven, team-theoretic one, focusing on issues of coordination. Of course, there very likely are incentive issues that arise even in the communication/inventories choice. For example, if production is for inventory and if sales personnel (who presumably are the source of information about the distribution of demand) are rewarded on the value of sales and not charged for excess inventories,[11] then there will be incentives to overstate demand. Nevertheless, incentives are not the only phenomena

[11] Either directly or through making the accuracy of their forecasts an element in determining their compensation.

of first-order importance in designing organizations, nor should they be treated as such in economic analyses.

References

Aggarwal, Sumer C.: MRP, JIT, OPT, FMS? Making sense of production operating systems. *Harvard Business Review 63*, no. 5, 8–16, September–October 1985.

Blond, Anthony: *The Publishing Game.* Jonathan Cape, 1971.

Chin, Lim: Theory of the firm: Uncertainty and choice of experiments. *Journal of Economic Theory 24*, 328–361, 1981.

Dudley, Leonard & Lasserre, Pierre: Information as a substitute for inventories. Cahier 8643, Départment de Science Economique, Université de Montréal, 1986. Forthcoming in *European Economic Review.*

Galbraith, Jay: *Designing Complex Organizations.* Addison-Wesley Publishing Co., Reading, MA, 1973.

First version submitted July 1987;
final version received January 1988.