Texturing: Using Make-to-Order Production Layered on Make-to-Stock Production to Hedge Against Uncertainty

Kyle Cattani
The Kenan-Flagler Business School
UNC Chapel Hill
Chapel Hill, NC 27599-3490
(919) 962-3273   kyle_cattani@unc.edu

Ely Dahan
Anderson School at UCLA
110 Westwood Plaza, B-514
Los Angeles, CA 90095
(310) 206-4170   ely.dahan@anderson.ucla.edu

Glen M. Schmidt
The McDonough School of Business
Georgetown University
Washington, DC 20057
(202) 687-4486   schmidtg@msb.edu

September 17, 2003

Texturing:

Using Make-to-Order Production Layered on Make-to-Stock Production to Hedge Against Uncertainty

Abstract

Consider a firm facing choosing between a strategy to pre-build a standard product using efficient capacity in advance of a single uncertain demand event (we refer to this as make-to-stock, MTS), or a strategy to acquire more expensive flexible capacity that can produce after observing the demand event (make-to-order, MTO). We identify the conditions under which strict MTS, strict MTO, or a dual approach utilizing both MTS and MTO are appropriate, and determine the efficient and flexible capacity requirements for each case. We refer to the dual approach as *texturing*: the firm pre-builds via MTS, and adds a layer of MTO output if needed. We extend the model to the case involving two standard products where the flexible capacity can fill demand for either. We explore how the fraction of MTO capacity varies as a function of capacity and obsolescence costs. We find that the firm may actually acquire more capacity than if the MTO capacity was dedicated to only one product. A texturing strategy is most attractive when obsolescence costs are high, when flexible production costs are predominantly variable (rather than fixed), and when the premium for flexible production is low.

Keywords: Operations Strategy, Manufacturing Strategy, Make-to-order, Make-to-stock, Capacity Management, Mass Production, Flexible Manufacturing

Texturing:
Using Make-to-order Production Layered on Make-to-Stock Production to Hedge Against Uncertainty

1. Introduction

Consider the idea of accurate response, as presented by Fisher and Raman (1996). Early in the production season of a fashion good, prior to demand realization, capacity is dedicated to products based on a forecast. Effectively production is make-to-stock (MTS). Later in the production season, demand is essentially known and production capacity can be used to respond to actual market demand. Effectively production is make-to-order (MTO). Fisher and Raman focus on determining the production sequence given a set production capacity (i.e., they establish which products should be produced at each stage of the season). In this paper we focus on the capacity decision. In particular, we seek insight into the appropriate amount of efficient, less expensive MTS production to have available prior to realizing demand and the appropriate amount of flexible, more expensive MTO product to have available for production in response to realized demand. We develop a simple single-period model to determine the optimal MTS and MTO capacity levels and the resulting average costs for first one product, and then extend the model to consider two products.

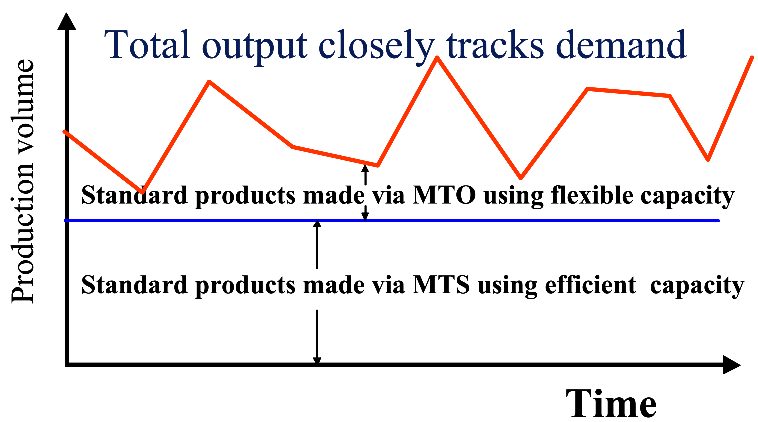
Alternately, consider a semiconductor manufacturer who holds two types of capacity. The manufacturer holds efficient capacity (possibly offshore) that produces the bulk of its goods in MTS fashion, but supplements this capacity with more expensive but also more flexible capacity (perhaps domestically situated) that can be used to fill "excess" demand when orders come in at a higher-than-expected level. The product is a standard item but will become obsolete quickly such that the plant has essentially one chance to set the production volume.

We consider these examples to illustrate dual, or hybrid, MTS – MTO approaches a firm can use to manage the trade-off between production costs and overage/underage costs. The dual strategy uses some efficient capacity operating in MTS mode to reduce production cost at the expense of overage/underage costs, while also using some flexible MTO production to reduce overage/underage costs but at the expense of production cost. In this paper we seek insights into the conditions under which a firm might adopt a dual approach, in a situation where demand exists solely for standard products.

MTO production using flexible capacity is often associated with the ability to make a wide variety of products on-demand. Such capabilities are needed when customized products are offered, for example. But when market demand is primarily for standard products, a firm might traditionally view this as a situation calling for efficient capacity and MTS production. (In the current paper, MTS production is always associated with efficient capacity and vice-versa. If the firm uses only efficient capacity we will refer to this as a mass production strategy.) Indeed, we find mass production to be preferred in our single period model, assuming there is a significant cost advantage for efficient resources. However, if the penalty associated with flexible capacity is less imposing, then we find the firm may want to consider a dual MTS – MTO approach we call *texturing*.

With the texturing strategy, the firm's primary production comes from efficient capacity and its resulting MTS production. But since MTS has the disadvantage of not exactly matching demand in the period, the firm restricts its efficient capacity (its MTS output) to avoid building excess stock of the standard product. Concurrently, it holds some separate flexible capacity that it uses in MTO fashion only if it realizes demand exceeding MTS capacity. The result is that efficient capacity is fully utilized, while flexible capacity is used only if there is sufficient demand. If our model was repeated as a series of independent and identical single-period problems, then varying levels of flexible capacity (MTO production) would be used in each period based on its specific demand realization. This would yield a total production level that is bumpy (it varies from period to period), but that closely matches the bumpy demand (see Figure 1), analogous to a textured (bumpy) wall or ceiling surface. Note that with texturing, the firm effectively operates two different production resources, efficient and flexible.

Figure 1 Texturing Yields a Bumpy Output Stream of Standard Products, From Two Production Resources.



In this paper, we first consider the case where the firm produces only a single standard product[1] in a single period. We find newsvendor-type results describing the optimal level of capacity the firm should hold, and offer an interpretation of the costs of underage and overage to

[1] Although we will continue to call this a single product, from a higher–level perspective within the firm we might also think of this as a single product line such as desktop computers, rather than as a single product. In other words, within a single product line the firm may make variations around a basic configuration, but the demand is treated as though it were only one product. Demand within the product line may be considered to be sufficiently *substitutable* to be treated as one product, or inventory might be reconfigured inexpensively after demand is known.

gain insight into the problem. We then extend this to the case of two products.[2] When offering two products, we assume the firm establishes separate efficient MTS capacities for each product, while the flexible MTO capacity can be used for either product. Again, we find newsvendor-type results.

We investigate the factors that determine when the dual strategy of texturing is preferred to mass production (strict MTS). We find that texturing is more attractive when obsolescence costs are high (i.e., when salvage value is low), which raises the upper bound on MTO cost for which a dual strategy applies. Also, as suggested earlier, texturing becomes more desirable as MTO production costs decrease. For example, initiatives such as flexible manufacturing, more efficient shop floor control resulting from more intense information processing, and information sharing up and down the supply chain may lead to more favorable MTO production costs. Thus future advances in technology, as they reduce MTO production cost and lead time, would seem to favor a shift to the dual strategy or even to strict MTO production (strict MTO becomes optimal as the cost penalty for MTO goes to zero).

With the two-product case, flexible capacity yields higher profits through demand pooling. Generally, pooling benefits result in lower resource requirements (such as holding lesser inventory). Interestingly, however, we find the firm may actually *increase* its overall capacity resource level when the flexible MTO capacity can fill demands for two-products, as compared to a situation where MTO capacity applies only to one product. While others have noted that pooling can lead to increased inventory (see Gerchak and Mossman (1992), for example), our

[2] Similar to the previous footnote, we can think of two products as representing two product lines, such as desktop computers and laptop computers, that share a common capacity constraint. We assume unique components and processes within the product line not to be a relevant constraint in our setting.

framework offers a specific context in which this result is observed. The flexible resource becomes more valuable in the two-product case, so the firm chooses to hold more of it.

The marketplace offers some examples of firms that have adopted the dual strategy in concept, if not directly as we model it here. For example, in the personal computer (PC) industry, some manufacturers have begun making computers in the MTO fashion (IBM, Gateway, and HP-Compaq), while still producing most of their computers using a traditional MTS approach. At the same time, this dual strategy does not appear to be universally preferred, or even necessarily the future dominant approach: Dell Computer generally is perceived as producing exclusively in the MTO fashion, while E-Machines, a low-price competitor, grew rapidly using the MTS approach, as noted by Hamilton (1999). We discuss some of the possible reasons we might observe these differences.

The paper is organized as follows. In § 2, we review related research. In § 3, we introduce the model and develop the results for the case of a single product. In § 4, we extend the model to two products. In § 5, we develop examples for hypothetical firms A, S, and C, that might abstractly represent firms in the automotive, semiconductor, and computer industries, respectively. We offer a discussion of results and conclusion in § 6.

2. Related Research

Our research relates to a broad stream that determines how best to supply products at low cost in the face of uncertain demand. Our approach is to assume, for a given product, that the firm may produce some units as MTS and others as MTO, but that any *given* unit follows either the MTS or the MTO route.

The notion of texturing contributes to the body of research that demonstrates the benefit of flexible resources. Eynan and Rosenblatt (1995) study the trade off between lower-cost MTS production and higher-cost MTO (or assemble-to–order, ATO) production to produce a single product. Our single-product model (§ 3) translates their result from an inventory decision into a capacity setting. In section 4, we extend the model to a two-product case.

Our two-product model (§ 4) relates to that of finding the optimal mix of less costly, dedicated capacity and more expensive, flexible capacity, as modeled by Van Mieghem (1998). In his model the firm decides between production using only dedicated capacity, only flexible capacity, or a mix of both, as determined by the marginal cost of flexible capacity. All production is MTO, in that capacities are allocated to products after demand is observed. Van Mieghem finds that it may be advantageous to invest in more expensive flexible resources even with perfectly positively correlated product demands. We also find support for expensive flexible resources but in a different framework with different timing: we assume that production can be MTO or MTS and that both fixed and variable costs are a function of the type of production, but MTS production must occur before demand is realized. Van Mieghem highlights the important role of price and cost mix differentials while our focus is on the benefit of flexible capacity to minimize overage and underage costs. Also, our models differ in regard to cost assumptions. Van Mieghem assumes fixed capacity costs are a function of the type of production while variable production costs are not; we assume the reserve.

Arreola-Risa and DeCroix (1998) study the optimality of a strict MTO policy versus a strict MTS policy for a company producing multiple heterogeneous products at a shared manufacturing facility. They model demands as independent Poisson processes with different arrival rates and derive optimality conditions for MTO versus MTS policies considering trade offs in inventory holding and backordering costs. They consider whether a specific product should be built MTO or MTS. Rajagopalan (2002) also considers the firm's portfolio of products using a model to determine which products should be made to order and which should be made to stock. We focus on solutions that consider dual use of MTO and MTS.

Cattani, et al. (2003) explore another dual MTS and MTO scenario for the case where there is customer demand for both standard and custom products. They focus on the benefits of achieving production smoothing (and thereby improving efficiency). The order stream for custom MTO products is assumed to be bumpy, so any remaining capacity is filled with production of standard items produced as MTS. They refer to this strategy as spackling.

3. Optimal Make–to–Stock and Make–to–Order Capacities for a Single Product

We consider a single period and in this section assume the firm makes a single standard product. Before observing demand the firm decides how much capacity to acquire, given that capacity comes in two types: efficient capacity that produces in MTS fashion, and capacity that is flexible in that it can produce in MTO fashion after demand is realized. Effectively, the firm chooses between three strategies: either it mass produces, acquiring only efficient capacity and producing only in MTS fashion (the level of flexible capacity is set to zero), or it acquires only flexible capacity and produces in MTO fashion, or it chooses a dual strategy we call texturing. With texturing, it acquires a strictly positive level of both capacity types, producing a base level of MTS output using the efficient resources, and then, after observing demand, producing up to capacity to satisfy any unfilled demand. Thus the total output varies with the demand realization (it is bumpy, or textured). The MTO capacity might be separated physically from the MTS capacity, in a domestic factory versus an overseas plant, or the capacities might be separated in time, such as in Fisher and Raman (1996), where the early use of the capacity is MTS and later use is MTO.

Demand is characterized as a non-negative, continuous random variable ***X*** with distribution $F(\cdot)$ and probability density $f(\cdot)$, and represents demand for one period (e.g., the season). The firm maximizes its expected profit by setting its efficient and flexible capacities, denoted by K_E and K_F, respectively. Total capacity is denoted by $K = K_E + K_F$. The firm produces K_E units before observing demand. If demand exceeds K_E, the firm fills this additional demand with output from the flexible capacity, up to its capacity level, K_F. Variable (per unit) costs are assumed to be constant over the volume of interest for each production type. The variable costs of producing one unit using efficient and flexible capacities are c_E and c_F, respectively. The selling price is p. For unmet demand, we assume lost sales. At the end of the period, any leftover units are salvaged at s. The model examines the ability of efficient and flexible production to deal with demand volume variability. These assumptions imply that the firm is not altering demand, or increasing product price, by initiating flexible production. Each type of capacity incurs a fixed cost of θ_i per unit per period with $i \in \{E,F\}$ for efficient and flexible capacities. The cost of capacity per period is thus $\theta_i K_i$ for each type.[3]

We make the following assumptions:

(A 1) $c_F, c_E, \theta_F, \theta_E > 0$. All production costs are greater than zero.

(A 2) $\theta_E + c_E > s$, $c_F > s$. Salvage value is less than the cost of production. There is no incentive to produce solely for the salvage market.

---

[3] Since efficient production is produced to capacity once capacity is decided, each unit of capacity will incur both the fixed cost θ_E as well as the variable cost c_E, so these costs could be combined.

(A 3) $c_F + \theta_F < p$ and $c_E + \theta_E < p$: We avoid trivial cases where the firm would never invest in MTS or MTO capacity, respectively.

In preparation for Theorem 1, define Δ as follows (interpretation follows the theorem):

$$\Delta \equiv \left(p-\left(c_F+\theta_F\right)\right)\left(c_F-s\right)/\left(p-c_F\right). \tag{1}$$

Theorem 1 gives the optimal efficient capacity K_E^* and the optimal flexible capacity K_F^*.

Theorem 1. The optimal capacities K_E^* and K_F^* are as follows:

Case A (strict MTO production): *If* $(c_F + \theta_F) - (c_E + \theta_E) \le 0$, *then*	$K_E^* = 0$ *and* $K_F^* = F^{-1}\left[\left(p-c_F-\theta_F\right)/\left(p-c_F\right)\right]$.	*(2)*
Case B (textured production): *If* $0 < (c_F + \theta_F) - (c_E + \theta_E) < \Delta$, *then*	$K_E^* = F^{-1}\left[((c_F + \theta_F) - (c_E + \theta_E)) / (c_F - s)\right]$, *and* $K_F^* = K^* - K_E^* = F^{-1}\left[\left(p-c_F-\theta_F\right)/\left(p-c_F\right)\right] - K_E^*$.	*(3)* *(4)*
Case C (mass production): *If* $\Delta \le (c_F + \theta_F) - (c_E + \theta_E)$, *then*	$K_E^* = F^{-1}\left[\left(p-c_E-\theta_E\right)/\left(p-s\right)\right]$ *and* $K_F^* = 0$.	*(5)*

All proofs are given in the Appendix.

Theorem 1 shows that the firm's optimal capacities K_E^* and K_F^*, depend on the cost premium for flexible production, $((c_F + \theta_F) - (c_E + \theta_E))$. Case A confirms that if the cost premium is nonexistent (if flexible production is no more expensive), then flexible production is unambiguously preferred and efficient production is not used, since this eliminates the cost of possible overstocks without adding any other costs. In this case the optimal flexible capacity K_F^* is given by equation (2), which is a critical fractile solution to a newsvendor–type problem. Here, the cost of underage is the cost of not building enough flexible capacity, $p - c_F - \theta_F$. The overage cost when flexible capacity exceeds demand is simply the marginal capacity cost θ_F.

Consider next Case B, the case of texturing, where $K_E^* > 0$ and $K_F^* > 0$. This case applies when the cost premium for flexible production is "not too high" (Δ defines this upper limit). Again, the results reflect newsvendor–type solutions. The flexible capacity as given in (4) is derived by first calculating the total capacity, K^*, and then subtracting out the efficient capacity, K_E^*, found by equation (3). In calculating the total capacity K^*, the cost of underage is the lost revenue p, minus the marginal cost $c_F + \theta_F$, since the last unit built will be using flexible capacity. The cost of overage is simply the marginal cost of the capacity itself, θ_F. K_E^*, as given by equation (3), is determined through a critical fractile where a shortfall of efficient production incurs an underage loss of $((c_F + \theta_F) - (c_E + \theta_E))$ per unit the firm spent extra by acquiring flexible production instead of the efficient type, and excess efficient production incurs an overage cost of $(c_E + \theta_E) - s - \theta_F$ per unit, since the firm lost $(c_E + \theta_E) - s$ by acquiring too much efficient capacity when it would have only lost θ_F had the capacity been of the flexible type.

Mass production (strict MTS using only efficient capacity) is unambiguously preferred (Case C) if the cost penalty for flexible production $(c_F + \theta_F) - (c_E + \theta_E)$ is sufficiently large, that is, if it exceeds Δ. This case arises, for example, when known and steady material flows facilitate fewer setups and less complexity, providing a significant cost advantage versus flexible production. Here we find $K_F^* = 0$, while (5) gives the optimal efficient capacity K_E^*. The cost of underage is the cost of insufficient efficient capacity, a lost profit per unit of $p - c_E - \theta_E$. The overage cost, incurred when efficient capacity (and therefore production) exceeds demand, is the capacity cost, minus the salvage value, $\theta_E + c_E - s$. We discuss further managerial implications after presenting an example.

3.1. *Example*

We explore comparative statics through an example where we assume demand ~U[0,100], p = \$100, $\theta_E = \theta_F = \$10$, c_E = \$65, and c_F = \$70. In Figure 2 we observe that for lower salvage value, s (i.e., higher obsolescence risk), the texturing model is beneficial. In such cases, it pays to texture with more costly flexible capacity in order to reduce expected overages. In this example, texturing is preferred if salvage value is less than \$62, non-textured efficient (mass) production is preferred otherwise.

Figure 2: The Effect of Salvage Value on the Texturing Model

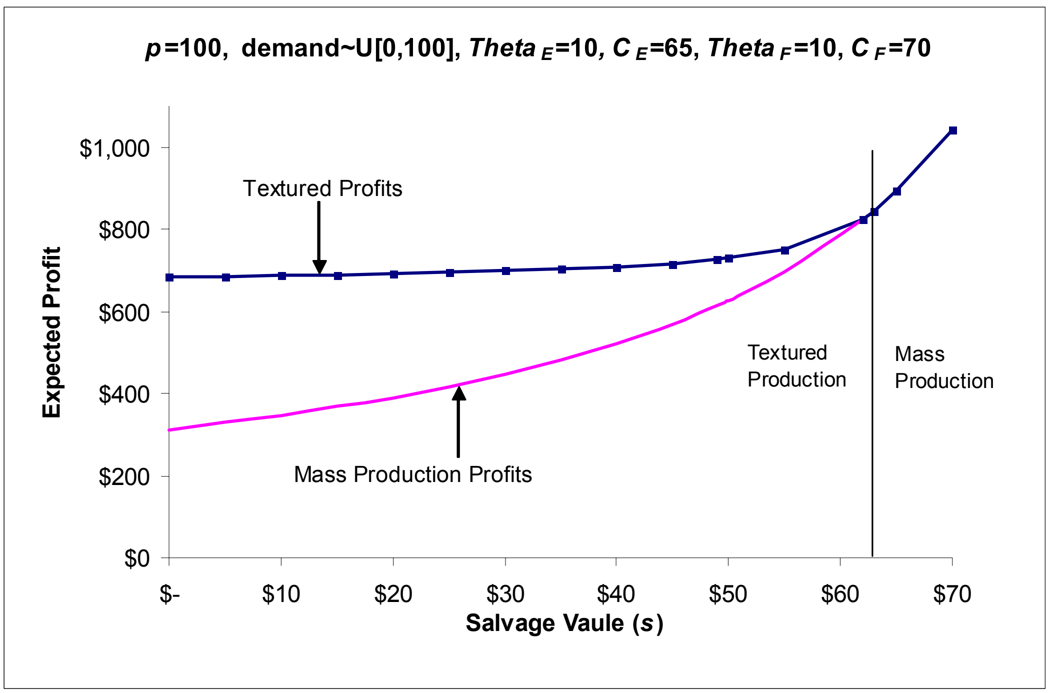
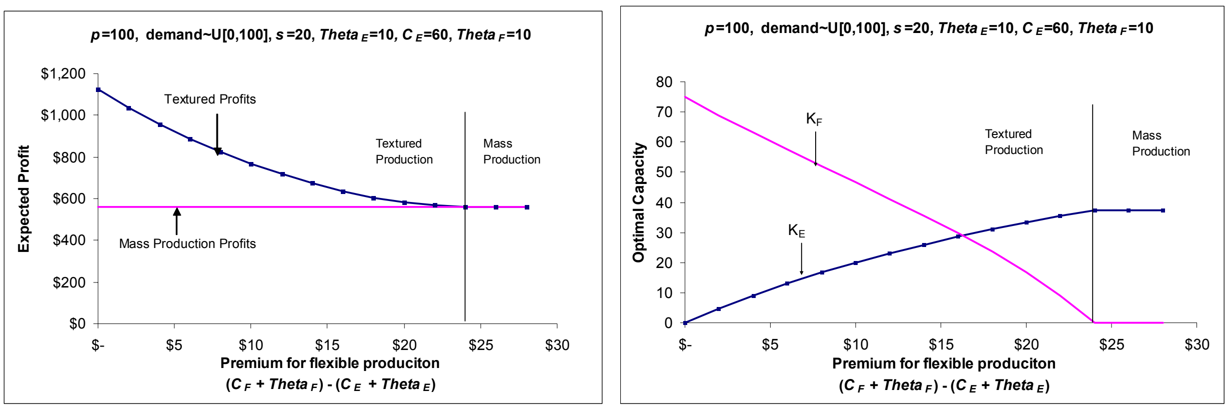


Figure 3 shows the optimal expected profits and capacities as a function of the cost premium for flexible production ((c_F + θ_F) – (c_E+ θ_E)). When flexible MTO production is sufficiently cost competitive with efficient MTS production (i.e., the premium is relatively low), made possible perhaps through new technology or supply chain management, the optimal MTO production increases and the optimal MTS production capacity decreases. In Figure 3, texturing is optimal for premiums up to $24, and mass production is optimal for premiums greater than $24. (Strict use of flexible production in this example is optimal when the cost premium is 0, or negative.)

Figure 3: The Effect of the Premium for Flexible Capacity



From our simple single-product model, we gain insights at a very basic level regarding some trends toward flexible production. Consider the computer industry. We hypothesize that the obsolescence cost (and consequently $(c_E+\theta_E)-s$) is relatively less for a producer of standard units such as E-Machines than for Dell, which focuses on leading-edge machines. (Lower-end products are presumably less sensitive to new product introductions.) $(c_F+\theta_F)-(c_E+\theta_E) > \Delta$ is equivalent to $c_F > p-\theta_F\left(c_P-s\right)/\left(c_E+\theta_E-s\right)$, and from Theorem 1 such a cost structure would lead to choosing efficient production, rather than textured production, if the unit cost of flexible production is sufficiently large. But this threshold should be relatively low for E-Machines.

One reason that PC firms may be moving to last minute production is that new technology or supply chain management techniques decrease $(c_F+\theta_F)-(c_E+\theta_E)$, the cost premium for MTO production. As production technology improves, and MTO production becomes more cost competitive with MTS production, the critical fractile representing the optimal MTS production capacity decreases.

With decreasing salvage value s, corresponding to increasing obsolescence costs, the critical fractile again decreases, favoring MTO production. For example, when the Pentium processor was introduced, HP-Compaq and IBM ended production of the 80486 models with excess inventory that had to be sold at fire-sale prices. These types of events illustrate the magnitude of obsolescence cost and provide a motivation for the computer companies to move to (at least some) MTO production.

Considering E-Machines, the new low-cost MTS entrant in the PC industry, we hypothesize that their obsolescence cost is relatively less as compared to Dell, who focuses on leading-edge machines (lower-end products are presumably less sensitive to new product introductions). This suggests that $(c_F - s)$ is relatively smaller for E-Machines, pushing them toward MTS production, while $(c_F - s)$ is relatively larger for Dell, making MTO production more attractive. Also, since this is a competitive market segment and E-Machines subcontracts its production, we hypothesize that $(p - c_F - \theta_F) / (p - c_F)$ is relatively small. (A significant percentage of the margin is paid to reserve capacity so θ_F is "large".) For MTO to be viable, E-Machine's cost premium for MTO production, $\left(c_F + \theta_F\right) - \left(c_E + \theta_E\right)$, must be even "smaller" than it would need to be for Dell. Thus the lower-technology, lower-margin emphasis of E-Machines seems to support its focus on MTS capacity as compared to Dell.

4. Two–Product Case

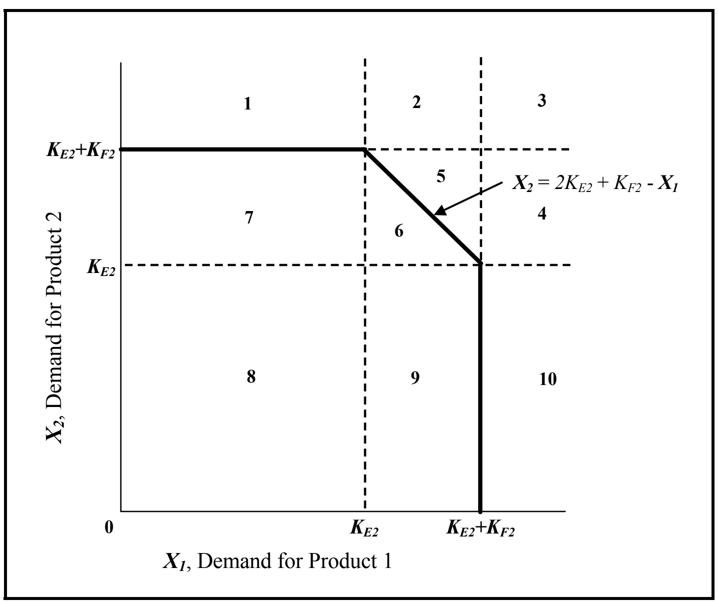
We have shown that the dual strategy may be optimal in the case of one product, as a means of managing demand volume uncertainty. We now analyze the firm's strategy with regard to MTS and MTO when offering multiple (two) products to gain insights into the effectiveness of the dual strategy to manage product mix issues.

The firm sets individual efficient MTS capacities for products 1 and 2, and a single flexible MTO capacity. The MTO capacity is flexible in that it can be used to fill demand of either product. Thus flexible capacity provides the benefit not only of responding to volume issues under stochastic demand but also to mix issues between the two products. In particular, if demand for one product is high while demand for the other is low, MTO capacity can be used to meet more of the demand for the high–demand product.

For simplicity we assume that prices, costs, and salvage values are the same for both products, and assume that demands for the two products are distributed independently and identically and that the marginal costs of capacity are the same for each type of production, $\theta_E = \theta_F = \theta$. We again assume (A 1) through (A 3). Thus, the problem is symmetric and the firm is assumed to set equal make-to-stock capacities for the two products. In the two-product case we use a subscript "2" and total capacity is $K_2 = K_{E2} + K_{E2} + K_{F2} = 2\, K_{E2} + K_{F2}$. Let $\boldsymbol{X_1}$ denote the demand for product 1, where $\boldsymbol{X_1}$ is a continuous non-negative random variable with cumulative distribution function $G(\cdot)$. Similarly, let $\boldsymbol{X_2}$ denote independent demand for product 2, also distributed $G(\cdot)$. Let $G_{12}(\cdot)$ denote the cumulative distribution function of the convolution of demand for products 1 and 2.

Expected profit $\Pi(K_{E2}, K_{F2})$ is determined through integration over the two–dimensional space for demand for products 1 and 2. As shown in Figure 4, there are 10 different regions to consider. In region 1, MTS production meets (and exceeds) product 1 demand, while demand for product 2 is not fully filled, in spite of using all K_{F2} units of MTO capacity for this product. In region 10, it is product 2 demand that is met while demand for product 1 is not fully filled. In regions 2-5 all MTS units are sold and all MTO capacity is used as well. In regions 6-9, all demand is met. In region 8, no MTO production is used and there are "leftovers" of both products (which are sold at salvage value). Regions 1, 7, 8, 9, and 10 result in some leftover make–to–stock units. Regions 6, 7, and 9 require use of some of the MTO capacity. In regions 1-5, and 10, all MTO capacity is used. Regions 2-5 require an allocation decision; since prices are the same the allocation does not affect expected profits. Without loss of generality, consider product 1 to have priority.

Figure 4: Regions of Potential Demand for Two–Products.



Lemma 1. *The expected profit function* $\Pi(K_{E2}, K_{F2})$ *is negative definite.*

Lemma 1 confirms that the profit function is well-behaved. Closed form solutions for the optimal capacities are not attainable, but it is possible to express optimal profits in terms of a newsvendor-type critical-fractile with an implicit solution as seen in the upcoming theorem.

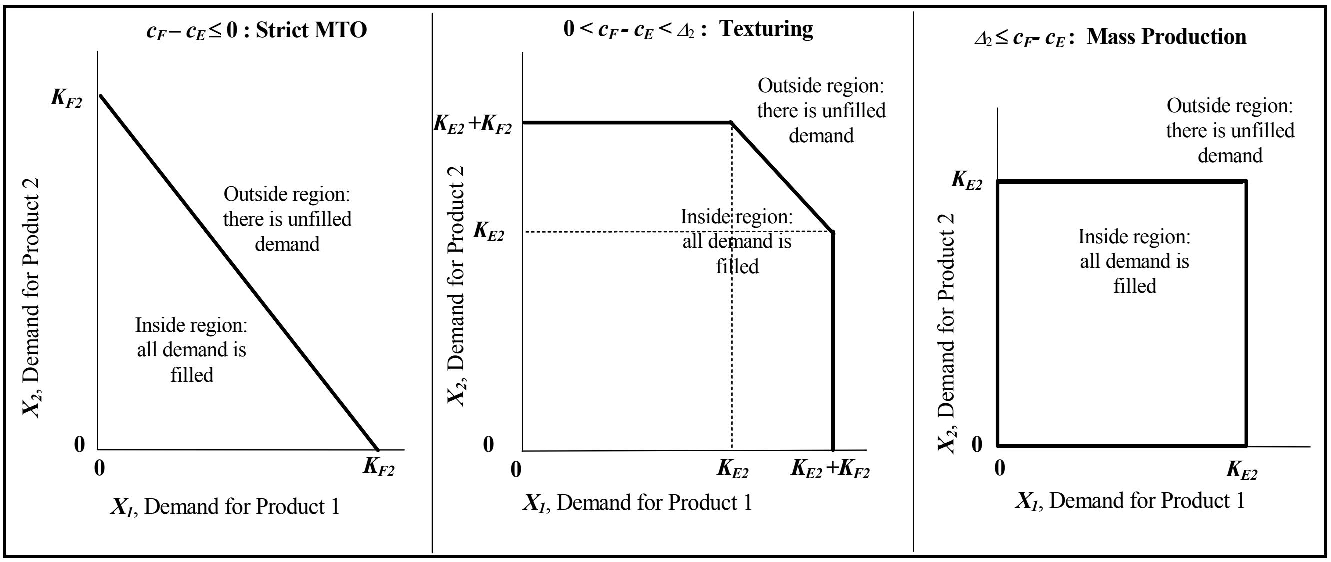
Theorem 2 identifies the optimal MTS capacity K_{E2}^*, the optimal MTO capacity K_{F2}^*, and the optimal total capacity K_2* for the two-product case. Define $I \equiv \int_{K_{E2}}^{K_{E2}+K_{F2}} \int_0^{2K_{E2}+K_{F2}-x_1} g(x_1) g(x_2) dx_2 dx_1$. Thus I represents the probability that demand falls within the combined regions 6 and 9 of Figure 4. Also define $\Delta_2 \equiv p - c_E - \frac{(p-s)^2 \theta}{(p-s)^2 - (p-c_E-\theta)^2}$. It will be shown that Δ_2 defines the upper limit for c_F relative to the dual strategy zone.

Theorem 2. *The optimal capacities K_{E2}^*, K_{F2}^* and K_2^* are the capacities that simultaneously solve the equations in Case A, B, or C, as the case applies:*

Case A (strict MTO production): *If $c_F - c_E \le 0$, then*	$K_{E2}^* = 0$ *and* $G_{12}(K_{F2}^*) = [(p - c_F - \theta) / (p - c_F)]$.	(6)
Case B (textured production): *If $0 < c_F - c_E < \Delta_2$, then*	$G(K_{E2}^*) = [(p - c_E - \theta) - (p - c_F) I) / (p - s)]$ *and* $G(K_{E2}^* + K_{F2}^*) = [p - c_F - \theta - (p - c_F) I] / [p - c_E - \theta - (p - c_F) I] - K_{E2}^*$.	(7) (8)
Case C (mass production): *If $\Delta_2 \le c_F - c_E$, then*	$G(K_2^*) = [(p - c_E - \theta) / (p - s)]$ *and* $K_{F2}^* = 0$.	(9)

Theorem 2 identifies three possible cases, as determined by the cost premium for flexible production, $c_F - c_E$ (see Figure 5 for a graphical representation). Similar to the one-product case (adding the simplification that $\theta_F = \theta_E = \theta$), if $c_F - c_E \le 0$, then all of the firm's production is MTO, with $K_{E2}^* = 0$ and K_{F2}^* as determined from equation (6). This again represents a traditional newsvendor solution. The form of the solution for Cases A and C is identical to the one–product case and is understood intuitively in the same manner. Case B (texturing) also has a critical fractile solution, but with more complex underage and overage costs that include the probability that demand falls in regions 6 and 9 of Figure 4. As $c_F - c_E$ increases from 0 up to Δ_2, the fraction of make-to-order production decreases and approaches zero. When this happens, equations (7) and (8) collapse to the traditional newsvendor solution with $K_{F2}^* = 0$ and K_{E2}^* as given by equation (9). Note that Δ_2 represents the lowest value of $c_F - c_E$ that leads to $K_{F2}* = 0$.

Figure 5: The cost premium $c_F - c_E$ establishes the production strategy.



4.1. *Example for Contrasting Firms*

Under the assumption that demand for each product is uniformly distributed over the interval [0, 100], we develop examples for three different firms, which we label as A, S, and C. We use these examples to illustrate how the MTS versus MTO capacity mix is affected by the fixed (per unit) capacity cost θ, the cost premium for MTO production (here we express it as the ratio c_F / c_E), and the obsolescence cost (the salvage value, s). To provide a common baseline, all examples use a sales price of p = 100. The characteristics of Firms A, S, and C are assumed to be as shown in Table 1.

Table 1: Characteristics of Firms A, S, and C.

Firm	Marginal fixed capacity cost, θ	Variable MTS cost, c_E	MTO cost premium ratio	Salvage value, s
A	Med: $\theta = 30$	Med: $c_E = 50$	Med: $c_F = 1.3\ c_E$	High: $s = 50$
S	High: $\theta = 50$	Low: $c_E = 15$	High: $c_F = 2\ c_E$	Low: $s = 10$
C	Low: $\theta = 10$	High: $c_E = 65$	Low: $c_F = 1.1\ c_E$	Low: $s = 20$

A key distinguishing factor of Firm A is its relatively higher salvage value: Firm A might produce automobiles rather than high tech products, because automobiles do not become obsolete as quickly. Key characteristics of Firm S include high fixed capacity cost and low salvage value: Firm S might be in the semiconductor manufacturing business, since fabrication facilities (fabs) are famously expensive, and microprocessors quickly lose their value when faster chips become available. The characteristics of Firm C are the same as those given previously in the example for the single-product case. As suggested earlier, we might position this firm in the computer industry, because of low fixed costs (high variable costs) and low salvage value. Our stylized single-period model cannot fully reflect all aspects of the suggested industries, but may provide some insights into how major differences in the general cost structures, and how differing obsolescence rates, might affect the MTS versus MTO decision making process.

We determine for each firm, using the parameter values in Table 1, optimal stocking levels and resulting expected profits for a base case of mass production where texturing is not an option (two products built in separate facilities using efficient capacities) versus our single-product model (1-prod Model) with texturing as an option versus our two-product model (2-prod Model) where texturing is again an option.

Figure 6: Optimal Capacities as Compared to the Case when Texturing is not an Option

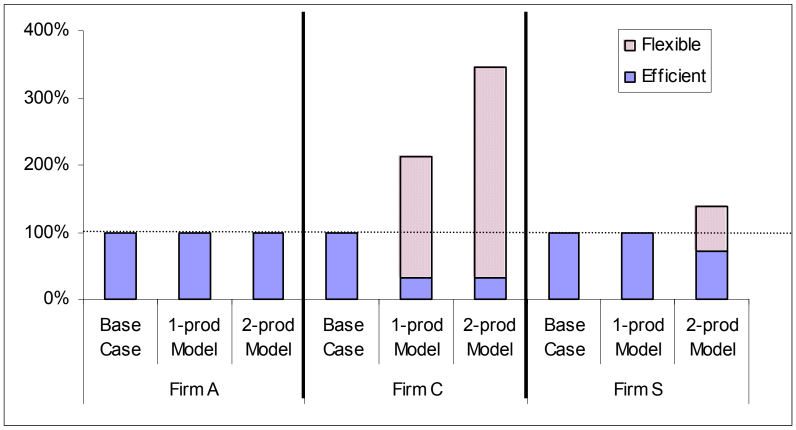
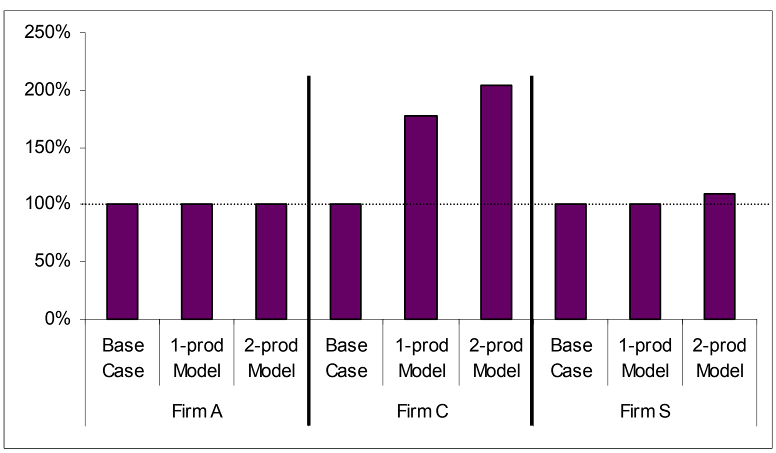


Figure 6 indicates that Firm A does not use MTO production in either the one-product or two-product cases, under these parameters. Firm C heavily relies on MTO production in both the one-product and two-product situations and Figure 7 confirms the benefits of doing so. Firm S employs MTO production only in the two-product case, benefiting less dramatically.

Figure 7: Total Expected Profit as Compared to the Case when Texturing is not an Option



Firm A, with its relatively higher salvage value and medium capacity and premium costs, does not benefit from texturing in this example, in either the one- or two-product case. Our results for Firm A concur with what we observe in the automobile industry, in that production is primarily MTS, with customers primarily buying cars off the dealer's lot rather than through special-order. However, as suggested earlier, automobile manufacturers have expressed a desire to switch to MTO production. To the extent that Firm A reflects the auto industry, our results suggest manufacturers need to bring the MTO cost down quite close to that of MTS in order for this switch to take materialize. Since the cost of MTO production includes lost customer goodwill created by an extended delivery lead-time, reducing lead time from today's level of weeks or even months to a matter of several days may be key to achieving the MTO goal.

In the two–product case, optimal expected profit is at least as high under shared capacity (one could always operate as though the capacity were not shared). Of the three firms, Firm C benefits the most from the opportunity to engage one or two-product texturing. The relatively low margins and low salvage values along with a relatively small premium for flexible manufacturing in this firm create an incentive to invest in flexible capacity that can avoid overages. In Figure 6 we note that Firm C utilizes mostly flexible capacity, given the option, and in Figure 7, Firm C has sharply higher expected profit in the one and two-product cases. In this example, the one-product texturing case allows Firm C to adjust to volume variances, and raises profits approximately 75% over the case where texturing is not allowed. The two-product case allows Firm C to adjust to mix variances, and raises profits an additional 15%. We conjecture that if more products were allowed (the *n*-product case), the benefits would continue to increase at a decreasing rate.

Firm S, with its high fixed cost of capacity, low salvage values, and relatively high premium for flexible production benefits from texturing only in the two-product case. Regarding the semiconductor industry, our model suggests that MTO production is more desirable when the fab is able to share capacity among multiple products. Consider American Semiconductor (a fictitious name disguising a real firm), as described by Pich and Harrison (1990). American produced wafers based on forecast (MTS). The fab was flexible to the extent that it could produce different families of related products, but product changeovers caused broken setups and disrupted maintenance schedules (suggesting a high cost premium for MTO). Finished wafers (each containing many die) were held in die bank inventory until firm orders were received. The case deals with the issue of using information systems to reduce the MTS inventory and, effectively, move the fab to more of a mixed MTS and MTO operation, as might be suggested by our model. (American also considers many other factors in this decision, such as dynamic scheduling and lead time issues.)

One might hypothesize that for any given cost parameters, less overall capacity is required in the two–product case than in the single-product case. Pooling of demands is generally associated with lower inventory levels, and in this situation the inventory is inventory of capacity. Our example shows the hypothesis to be false for Firms C and S.

5. Discussion and Summary

The MTS versus MTO choice typically has been viewed as an either/or decision – either MTS or MTO. The choice has typically been for MTS (mass production) when standard items are involved. However, we show that a mix of MTS/MTO (texturing) can be useful in certain situations.

There are other advantages and implications of a dual strategy approach that we do not address. For example, the National Bicycle Industrial Company, as described by Fisher (1994), uses MTO production for lower-volume, up-scale bikes, while sticking with MTS production for the more basic models. Thus the dual strategy expands their product line toward the upper end, and has significant implications on both overall product variety and product demand. A customer might be willing to pay more for an MTO product because such a product more precisely meets her needs, as opposed to off–the–shelf products that may contain unwanted features or lack other desired features. In contrast, our results apply in the absence of any market expansion or price premiums that might accrue from MTO production, since we assume the price and demand are unaffected by the choice of MTS or MTO. Also, we do not explicitly consider potential negative market aspects of MTO production, such as the possibility that customers may become impatient in waiting for MTO products. Possible extensions of our model might model explicitly these effects.

Our investigations involving Firms A, S, and C offer a glimpse into the possible implications of changes in parameters such as the cost premium for MTO production, fixed capacity cost, and obsolescence cost. Specifically, we showed an example (Firm A) where MTO capacity was not particularly useful in either the one or two-product cases, a situation (Firm S) where MTO capacity was valuable in only the two-product case, and an example (Firm C) where MTO capacity was utilized in both cases. While we found some similarities between these firms and characteristics within the automotive, semiconductor, and computer industries, additional questions remain. For example, when considering two products, what is the impact of correlation between demands? We presume benefits of the flexible MTO capacity would be less pronounced as the correlation becomes positive between products, and even more pronounced with negative correlation between products, but desire more clarity in this regard. Also, how are our results affected by the *magnitude* of differences in features between the two products? For example, in the semiconductor industry, what is the implication of having the flexibility to make, say, the Pentium® IV processor with multiple processing speeds, versus the flexibility to make the Pentium® III along with the Pentium® II, versus the flexibility to make microprocessors along with DRAM?

As the above discussion suggests, the firm has many factors to consider in its decision of MTS versus MTO. We focus on only one of these, exploring how the choice of MTS capacity and MTO capacity might help the firm deal with demand uncertainty. Toward this end, we provide some insights into the impact of product obsolescence costs, capacity costs, and the cost premium of flexible MTO production. Thus our model offers some incremental understanding of the factors pushing the firm toward strict MTS, strict MTO, or the dual strategy approach, and is intended to spur further investigation in this regard.

Appendix: Proofs

Proof of Theorem 1.

The firm maximizes expected profit $\Pi(K_E, K_F)$ by setting investments K_E and K_F, subject to the investments being nonnegative. The optimization problem is:

Max K_E, K_F

$$\Pi(K_E,K_F) = p\int_0^{K_E+K_F} xf(x)dx + p\left(K_E+K_F\right)\left(1-F\left(K_E+K_F\right)\right) + s\int_0^{K_E}(K_E-x)f(x)dx - c_E K_E$$
$$-c_F\int_{K_E}^{K_E+K_F}\left(x-K_E\right)f(x)dx - c_F K_F\left(1-F\left(K_E+K_F\right)\right) - \theta_E K_E - \theta_F K_F$$

Subject to: $K_E,\ K_F \geq 0$.

First Order Conditions:

The solution to the first order condition with respect to K_F is unique and is given by:

$\dfrac{\partial\Pi\left(K_E,K_F\right)}{\partial K_F} = \left(p-c_F\right)\left(1-F\left(K_F+K_E\right)\right) - \theta_F = 0$, which leads to:

$$F(K_F+K_E) = \frac{p-c_F-\theta_F}{p-c_F} \tag{10}$$

The solution to the first order condition with respect to K_s is unique and is given by:

$\dfrac{\partial\Pi\left(K_E,K_F\right)}{\partial K_E} = -c_E + p + \left(s-c_F\right)F\left(K_E\right) + \left(c_F-p\right)F\left(K_F+K_E\right) - \theta_E = 0$, which leads to:

$$F\left(K_E\right) = \frac{\left(p-c_E\right)-\left(p-c_F\right)F\left(K_F+K_E\right)-\theta_E}{c_F-s}. \tag{11}$$

Substituting (10) into (11), we find:

$$F(K_E) = ((c_F+\theta_F) - (c_E+\theta_E)) / (c_F - s), \tag{12}$$

which leads directly to equation (3). Substituting (12) back into (11) yields:

$$F(K_E+K_F) = [(p - c_F - \theta_F) / (p - c_F)], \tag{13}$$

from which (4) is readily derived. The constraint that $K_F \geq 0$ equates to $F(K_E) \leq F(K_E+K_F)$. Using (12) and (13), this equates to $(c_F + \theta_F) - (c_E + \theta_E) \leq (p - c_F - \theta_F)(c_F - s) / (p - c_F)$, or equivalently, $(c_F + \theta_F) - (c_E + \theta_E) \leq \Delta$. When $(c_F + \theta_F) - (c_E + \theta_E) = \Delta$, it follows that $F(K_E+K_F) = F(K_E)$ and thus $K_F = 0$. Having shown (3), (4), and the boundary condition, this completes the proof of Case B. Note that when $(c_F + \theta_F) = (c_E + \theta_E)$, equation (12) shows $F(K_E)$

= 0 and in this case equation (13) shows $F(K_F) = [(p - c_F - \theta_F) / (p - c_F)]$, leading to equation (2) and completing the proof of Case A.

The above analysis shows that when $\Delta < (c_F + \theta_F) - (c_E + \theta_E)$, the constraint $K_F \geq 0$ is binding, and thus $F(K_E{+}K_F) = F(K_E)$. Substituting $F(K_E{+}K_F) = F(K_E)$ into equation (11), and solving for $F(K_E)$ yields $F(K_E) = [(p - c_E - \theta_E) / (p - s)]$, confirming equation (5) and completing the proof of Case C.

Second Order Conditions:

The Hessian for the profit function with respect to K_F and K_E is:

$$H = \begin{bmatrix} -(c_F - s) f(K_E) - (p - c_F) f(K_E + K_F) & -(p - c_F) f(K_E + K_F) \\ -(p - c_F) f(K_E + K_F) & -(p - c_F) f(K_E + K_F) \end{bmatrix}$$ This is of the form:

$$H = \begin{bmatrix} -c - b & -b \\ -b & -b \end{bmatrix}$$ where $c = (c_F - s) f(K_E) > 0$, and $b = (p - c_F) f(K_E + K_F) \geq 0$.

The Hessian is negative definite in regions where $-c - b < 0$ and $|H| > 0$. Since $|H| = b$ c, and we know that $c > 0$, we conclude the Hessian is negative definite if $0 > -b$, i.e., if $0 > -(p - c_F) f(K_F + K_E)$. Thus, the Hessian is negative definite for all values of K_F and K_E, and a solution meeting the first order conditions and the constraint is globally optimal.

Proof of Lemma 1 and Theorem 2.

The firm's objective is to maximize expected profit, $\Pi(K_{E2}, K_{F2})$, as given by equation (14) below, subject to $K_{E2} \geq 0$ and $K_{F2} \geq 0$:

$$\begin{aligned}
\Pi(K_{E2}, K_{F2}) = {} & p\int_0^{K_{E2}}\int_0^{K_{E2}+K_{F2}} (x_1 + x_2) g(x_1) g(x_2)\, dx_2 dx_1 + p\int_{K_{E2}}^{K_{E2}+K_{F2}}\int_0^{2K_{E2}+K_{F2}-x_1} (x_1 + x_2) g(x_1) g(x_2)\, dx_2 dx_1 \\
& + 2p\int_0^{K_{E2}}\int_{K_{E2}+K_{F2}}^{\infty} (x_1 + K_{E2} + K_{F2}) g(x_1) g(x_2)\, dx_2 dx_1 + p\int_{K_{E2}+K_{F2}}^{\infty}\int_{K_{E2}}^{\infty} (2K_{E2} + K_{F2}) g(x_1) g(x_2)\, d \\
& + p\int_{K_{E2}}^{K_{E2}+K_{F2}}\int_{2K_{E2}+K_{F2}-x_1}^{\infty} (2K_{E2} + K_{F2}) g(x_1) g(x_2)\, dx_2 dx_1 + s\int_0^{K_{E2}}\int_0^{K_{E2}} (2K_{E2} - x_1 - x_2) g(x_1) g(x_2) \\
& + 2s\int_0^{K_{E2}}\int_{K_{E2}}^{\infty} (K_{E2} - x_1) g(x_1) g(x_2)\, dx_2 dx_1 - 2c_F\int_0^{K_{E2}}\int_{K_{E2}}^{K_{E2}+K_{F2}} (x_2 - K_{E2}) g(x_1) g(x_2)\, dx_2 dx_1 \\
& - c_F K_{F2}\left\{(1 - G(K_{E2} + K_{F2}))(1 + G(K_{E2})) + \int_{K_{E2}}^{K_{E2}+K_{F2}}\int_{2K_{E2}+K_{F2}-x_1}^{\infty} g(x_1) g(x_2)\, dx_2 dx_1\right\} \\
& - c_F\int_{K_{E2}}^{K_{E2}+K_{F2}}\int_0^{2K_{E2}+K_{F2}-x_1} (x_1 + x_2 - 2K_{E2}) g(x_1) g(x_2)\, dx_2 dx_1 - 2K_{E2}c_E - \theta(2K_{E2} + K_{F2})
\end{aligned} \tag{14}$$

First Order Conditions:

The first order conditions for optimality are as follows:

$$\begin{aligned}\frac{\partial\Pi\left(K_{E2},K_{F2}\right)}{\partial K_{E2}} &= p\iint\limits_{\text{regions 1-5}} g\left(x_1\right)g\left(x_2\right)dx_2dx_1 + p\iint\limits_{2-5,10} g\left(x_1\right)g\left(x_2\right)dx_2dx_1 \\ &+s\iint\limits_{8,9,10} g\left(x_1\right)g\left(x_2\right)dx_2dx_1 + s\iint\limits_{1,7,8} g\left(x_1\right)g\left(x_2\right)dx_2dx_1 \\ &+c_F\iint\limits_{6,9} g\left(x_1\right)g\left(x_2\right)dx_2dx_1 + c_F\iint\limits_{6,7} g\left(x_1\right)g\left(x_2\right)dx_2dx_1 - 2c_E - 4\theta \\ &= 0\end{aligned} \tag{15}$$

and

$$\frac{\partial\Pi\left(K_{E2},K_{F2}\right)}{\partial K_{F2}} = \left(p-c_F\right)\iint\limits_{1-5,10} g\left(x_1\right)g\left(x_2\right)dx_2dx_1 - \theta = 0 \tag{16}$$

Equations (15) and (16) can be used to develop the following system of equations; (15) is used to develop (17) and (16) is used to develop (18).

$$G\left(K_{E2}\right) = \frac{\left(p-c_E-2\theta\right)-\left(p-c_F\right)I}{\left(p-s\right)} \tag{17}$$

$$G(K_{E2})\ G(K_{E2}+K_{F2}) = 1 - I - \theta\ /\ (p-c_F) \tag{18}$$

where $I = \int_{K_{E2}}^{K_{E2}+K_{F2}}\int_0^{2K_{E2}+K_{F2}-x_1} g\left(x_2\right)g\left(x_1\right)dx_2dx_1$, the probability of demand falling within the zone of combined regions 6 and 9.

Subtracting (18) from (17) yields (19), which no longer contains *I*:

$$G\left(K_2\right) = \frac{\left(c_F-c_E\right)}{\left(p-s\right)-\left(p-c_F\right)G\left(K_{E2}+K_{F2}\right)} \tag{19}$$

Equation (17) is used directly to determine equation (7). After substituting equation (17) into (18), the result is used to determine equation (8). Boundary conditions for $c_F - c_E$ (i.e., Δ_2) are determined to ensure that K_{E2}* and K_{F2}* are greater than 0.

Second Order Conditions:

The Hessian for the profit function, equation (14), with respect to K_{F2} and K_{E2} is:

$$H = \begin{bmatrix} -(p-c_F)\frac{\partial I}{\partial K_{F2}} - (p-c_F)G(K_{E2})g(K_{E2}+K_{F2}) & -2(p-c_F)\frac{\partial I}{\partial K_{F2}} \\ -2(p-c_F)\frac{\partial I}{\partial K_{F2}} & -2(p-s)g(K_{E2}) - 2(p-c_F)\frac{\partial I}{\partial K_{E2}} \end{bmatrix}$$ where

$$\frac{\partial I}{\partial K_{F2}} = \int_{K_{E2}}^{K_{E2}+K_{F2}} g(2K_{E2}+K_{F2}-x)g(x)dx + g(K_{E2}+K_{F2})G(K_{E2}),$$ and

$$\frac{\partial I}{\partial K_{E2}} = 2\int_{K_{E2}}^{K_{E2}+K_{F2}} g(2K_{E2}+K_{F2}-x)g(x)dx + g(K_{E2}+K_{F2})G(K_{E2}) - g(K_{E2})G(K_{E2}+K_{F2}).$$

Let:

$$a = (p-c_F) > 0,$$
$$b = G(K_{E2})g(K_{E2}+K_{F2}) \geq 0,$$
$$d = (p-s) > 0,$$
$$e = g(K_{E2}) > 0,$$
$$h = \int_{K_{E2}}^{K_{E2}+K_{F2}} g(2K_{E2}+K_{F2}-x)g(x)dx \geq 0, \text{ and}$$
$$i = G(K_{E2}+K_{F2}) > 0.$$

Now let $\alpha = -a(2b+h) \leq 0$, and $\beta = -\dfrac{(de-aei)a(2b+h)+a^2bh}{ab+(de-aei)} \leq 0$.

Thus, $\dfrac{\partial I}{\partial K_{F2}} = h+b$, and $\dfrac{\partial I}{\partial K_{E2}} = 2h+b-ei$. The Hessian can now be written as:

$$H = \begin{bmatrix} -a(h+b)-ab & -2a(h+b) \\ -2a(h+b) & -2de-2a(2h+b-ei) \end{bmatrix}.$$ The Hessian is negative definite in regions where $-a(h+b)-ab<0$ and $|H|>0$. Since $-a(h+b)-ab<0 \Leftrightarrow 0>-a(2b+h)$ and since $|H| = (de-aei)2a(2b+h)+2a^2bh$, the Hessian is negative semi–definite for all values of K_{F2} and K_{E2} when $0>-a(2b+h) = \alpha$ and $0 > -\dfrac{(de-aei)a(2b+h)+a^2bh}{ab+(de-aei)} = \beta$. Note that α and β are less than or equal to zero, since $de-aei>0$. Negative definiteness is assured, such that a solution meeting the first order conditions and the constraints is globally optimal.

References

1998. Mcjiters. The Economist. 347 70.

Arreola-Risa, A. and G. A. DeCroix. 1998. Make-to-order versus make-to-stock in a production-inventory system with general production times. IIE Transactions. 30 705-713.

Cattani, K. D., E. Dahan and G. Schmidt. 2003. Spackling: Smoothing make-to-order production of custom products with make-to-stock production of standard items. Under review at Management Science.

Eynan, A. and M. J. Rosenblatt. 1995. Assemble to order and assemble in advance in a single-period stochastic environment. Naval Research Logistics. 42 861-872.

Fisher, M. 1994. National bicycle industrial company case study. The Wharton School. University of Pennsylvania, Philadelphia, PA

Fisher, M. and A. Raman. 1996. Reducing the cost of demand uncertainty through accurate response to early sales. Operations Research. 44 (1) 87-99.

Gerchak, Y. and D. Mossman. 1992. On the effect of demand randomness on inventories and costs. Operations Research. Jul/Aug. 40 (4) 804-807.

Gibson, R. 1999. Mcdonald's 1st period net rose 11%, but troubles overseas rattle investors. The Wall Street Journal. (April 23, 1999)

Hamilton, D. P. 1999. Pcs for under $600 seize a chunk of the market. The Wall Street Journal. (March 26, 1999)

Pich, M. T. and M. Harrison. 1990. American semiconductor, inc. (a) service systems architecture. Graduate School of Business, Stanford University. Stanford, CA.

Rajagopalan, S. 2002. Make to order or make to stock: Model and application. Management Science. 48 241-256.

Van Mieghem, J. A. 1998. Investment strategies for flexible resources. Management Science. 44 (8) 1071-1078.