Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Market Segmentation, Advanced Demand Information, and Supply Chain Performance

Fangruo Chen,

Market Segmentation, Advanced Demand Information, and Supply Chain Performance

Fangruo Chen
Graduate School of Business, Columbia University, New York 10027
fc26@columbia.edu

A monopolist sells a single product to a market where the customers may be enticed to accept a delay as to when their orders are shipped. The enticement is a discounted price for the product. The market consists of several segments with different degrees of aversion to delays. The firm offers a price schedule under which the customers each self-select the price they pay and when their orders are to be shipped. When a customer agrees to wait, the firm gains advanced demand information that can be used to reduce its supply chain costs. This article shows how an optimal pricing-replenishment strategy that balances the costs due to discounted prices and the benefits due to advanced demand information can be determined.
(*Supply Chain Management; Market Segmentation; Pricing; Shipping Delays; Value of Information; Product Variety*)

1. Introduction

Suppose you are considering buying a book from Amazon.com. You check their Web site and find the book (it's hard to miss a title, given Amazon's huge collection). If the list price for the book is acceptable, you add it to your shopping cart. Before checkout, you are asked to choose from three shipping methods: standard shipping, second day air, or next day air. If you can wait a few days, you choose standard shipping; otherwise, you pick a faster method. After mulling over the three options, you select one and send the order to Amazon.com. The total price equals the book's list price plus the shipping fee. Assuming the book is in stock in Amazon.com's distribution center, you will receive it after a transit time. From the standpoint of the firm's operations (assuming a third party is responsible for transporting customer orders), the different shipping methods are identical: The distribution center strives to ship a customer order as soon as it is received.

Whereas some book buyers have an urge to obtain the book right away, others may be willing to delay the shipment of their orders *if* the list price is properly discounted.[1] This provides an opportunity for the firm to segment the market so that impatient customers pay the list price and have their orders shipped immediately, whereas the patient ones pay a discounted price and agree to postpone the shipment of their orders. When the firm receives an order that does not have to be shipped immediately, it gains advanced demand information, in other words, a preview of the future demand process. This information can be used to cut the firm's inventory. The optimal price discount should, of course, balance the revenues foregone with the value of the extracted information.

The objective of this paper is to study the benefits of the above segmentation strategy. We consider a model where a monopolist sells a single product to heterogeneous customers. The customers divide into

[1]The author recently bought three books from Amazon.com. They have been sitting on his desk for a month now.

1523-4614/01/0301/0053$05.00
1526-5498 electronic ISSN

segments exhibiting different degrees of aversion to waiting: The highest price they are willing to pay for the product, i.e., the reservation price, decreases as the delay increases, and different segments have different reservation-price functions. The firm owns a multi-stage supply chain that replenishes the finished-goods inventory. The objective is to make pricing and replenishment decisions to maximize the firm's long-run average profits. I show how this problem can be solved and present numerical examples that show that in some cases, the firm stands to gain substantially by allowing customers to self-select the price they pay and when their orders are to be shipped.

There is a growing body of research that addresses the value of information in managing supply chains. Closely related to the current research are Milgrom and Roberts (1988), Hariharan and Zipkin (1995), and Lovejoy and Whang (1995). Milgrom and Roberts show that inventory and demand information are substitutes for one another. In their model, additional demand information is learned via a market survey. In the model presented here, demand information is extracted through pricing and consumer self-selection. Hariharan and Zipkin demonstrate that customers' advanced warnings of orders are equivalent to reductions in replenishment leadtimes. They assume that every customer places an order l units of time ahead of the actual demand. The value of l is exogenously given in their model, whereas customers in this research have to be enticed to choose their own shipping delays. (Hariharan and Zipkin refer to l as the demand leadtime.) Lovejoy and Whang consider a model where each arriving order is first processed by an information processing system and then downloaded to a production system. Longer information processing times incur longer delays but provide more advanced warnings about the upcoming demand to the production function. They show how this trade-off can be made. Therefore, the advanced warnings are an internal decision for the firm, without any customer participation. The reader is referred to a recent book edited by Tayur, Ganeshan, and Magazine (1998) for a comprehensive review of this literature.

This research is also related to the vast literature on market segmentation, a marketing strategy that represents a rational and precise adjustment of product and marketing efforts to satisfy divergent customer demands (e.g., Smith 1956 and Haley 1968). This strategy has led to the unprecedented proliferation of product lines (Quelch and Kenny 1994) and, lately, mass customization, i.e., individually tailored products offered on a large scale (Pine 1993). Research on product variety is built on the notion that a product can be viewed as a bundle of characteristics or attributes, and consumers are interested in the product not for its own sake but because of the attributes it possesses (Lancaster 1979). For example, one drinks orange juice for vitamin C and, lately, calcium. Given divergent consumer preferences over different bundles of product attributes, the question is how a firm should position its products in the attribute space. Many papers have addressed this question from different angles, e.g., Hotelling (1929), Mussa and Rosen (1978), Hauser and Simmie (1981), Moorthy (1984), Eliashberg and Manrai (1989), Lancaster (1990), deGroote (1994), Nanda (1995), and Chen, Eliashberg, and Zipkin (1998). Whereas these authors typically focus on physical attributes (e.g., size, color, sweetness), this paper shows that a product's delivery schedule can also be a useful attribute for segmenting the market (see Ho and Tang (1998) for the latest developments in the area of product variety management).

Finally, there is an extensive literature on managing customers with heterogeneous costs of waiting in queueing environments, e.g., Mendelson and Whang (1990) and the references therein. A key distinction between the current paper and the queueing literature is that in this model, the delays experienced by the customers are actually used to reduce the costs in serving them (i.e., the inventory costs incurred by the supply chain).

The rest of the paper is organized as follows. Section 2 introduces the model. Section 3 characterizes several properties of an optimal pricing strategy. Section 4 shows how advanced demand information can lead to lower inventory costs. Section 5 describes algorithms for computing an optimal pricing strategy. Section 6 reports numerical examples. Section 7 concludes.

2. The Model

A firm sells a single product to customers. The selling process has two stages. At the first stage, potential customers visit the point of sales (e.g., Amazon.com's on-line catalogue), where they see a list price for the product. Those who find the list price acceptable enter the second stage; the rest leave. At the second stage, the customers are presented with a price schedule, which specifies a price discount for each of several possible shipping dates. Each customer chooses a shipping date, pays the corresponding price, and buys one unit of the product.

I focus on the second stage of selling. Let p_0 be the list price, which is fixed. Customers arrive at the second stage of the selling process according to a Poisson process with average rate λ_0. The price schedule is $\{(p_k, \tau_k)\}_{k=0}^{K}$, $K \geq 0$ integer, where p_k is the final price if a customer agrees to have his order shipped τ_k units of time after order placement, with $0 = \tau_0 < \tau_1 < \ldots < \tau_K$. Thus, if a customer wants immediate shipment of his order, he pays the list price, i.e., choosing (p_0, τ_0). The remaining pairs in the price schedule provide price discounts to customers who are willing to wait. Thus, $p_k < p_0$ for all $k > 0$. The first objective is to identify an optimal price schedule that maximizes the firm's long-run average profits.

Different customers have different costs of waiting. Suppose the customer population (entering the second stage of selling) consists of M segments or types. Let $u_m(\tau)$ be the maximum (or reservation) price that the customers in segment m are willing to pay for one unit of the product if their orders are shipped τ units of time after order placement, $m = 1, \ldots, M$. Assume that $u_m(\cdot)$ is decreasing, differentiable, convex with $u'_m(\tau) < u'_{m+1}(\tau)$ (< 0) for all τ, i.e., $u_{m+1}(\cdot)$ is flatter than $u_m(\cdot)$, $m = 1, \ldots, M - 1$. Thus, Segment 1 is uniformly more sensitive to waiting than Segment 2, Segment 2 more so than Segment 3, and so on. Assume $u_m(0) \geq p_0$ for $m = 1, \ldots, M$, which is true if the customers initially think that all orders are shipped immediately and enter the second stage of selling because their corresponding reservation price exceeds the list price.[2]

Note that $u_m(0) - u_m(\tau)$ represents the cost to a type-m customer for agreeing to wait τ units of time. The assumption that $u_m(\cdot)$ is convex implies that the cost of waiting is concave, i.e., the marginal cost of waiting is decreasing. This is true if, for example, the excitement about the product decreases over time after the order is placed. In this case, the marginal cost of waiting is very high in the first few days, while the excitement still lingers, and decreases as time goes by. Here are two examples that satisfy the above assumptions on the reservation prices:

$$u_m(\tau) = u_m(0) - \theta_m(1 - e^{-\tau}),\ m = 1, \ldots, M, \quad (1)$$

or

$$u_m(\tau) = u_m(0) - \theta_m \tau,\ m = 1, \ldots, M, \quad (2)$$

where θ_m are positive constants with $\theta_1 > \theta_2 > \ldots > \theta_M$. Obviously, to avoid negative $u_m(\tau)$, one must restrict to $u_m(0) - \theta_m \geq 0$ for all m in (1) and to $\tau \leq u_m(0)/\theta_m$ for all m in (2).

The surplus that a type-m customer derives from (p_k, τ_k) is $u_m(\tau_k) - p_k$. His objective is to choose a pair from the price schedule to maximize his surplus. Because $(p_0, \tau_0 = 0)$ is always available and $u_m(0) - p_0 \geq 0$ as assumed above, each customer is guaranteed a non-negative surplus. Thus, no customers are priced out of the market. We assume that type-m customers arrive according to a Poisson process with average rate λ_m, $m = 1, \ldots, M$. Thus, $\Sigma_{m=1}^{M} \lambda_m = \lambda_0$.

The firm owns an N-stage supply chain. Stage 1 is the final stocking point from which the product is shipped to customers, Stage 1 is replenished by Stage 2, Stage 2 by Stage 3, etc., and Stage N by an outside supplier with ample stock. The transit time from one stage to the next is constant. Customer orders are sequenced according to their shipping dates, not the dates on which the orders are placed. This sequence determines which order is shipped next in a first-come, first-served manner. If the firm cannot ship an order on the date chosen by the customer because the product is out of stock (at Stage 1), I assume that the order is backlogged. The backlogged orders are shipped as soon as inventory becomes available. In this case, the firm incurs a goodwill loss (or backorder cost). In addition, the firm incurs holding costs for inventories

[2]Of course, the entire price schedule may be made available to all potential customers when they first arrive at the point of sales. The risk is that this may create confusion when customers are trying to decide which product to buy among many others in, e.g., an on-line catalog.

held in the supply chain and variable costs for every unit sold. For $n = 1, \ldots, N$, define:

L_n = transit time or leadtime from stage $n + 1$ to stage n, with stage $N + 1$ being the outside supplier.
H_n = holding cost rate at stage n (per unit per unit of time).
h_n = echelon holding cost rate at stage n.
= $H_n - H_{n+1} > 0$, with $H_{N+1} = 0$.
b = backorder cost rate (at Stage 1, per unit per unit of time).
c = unit variable cost (e.g., procurement, production, and transportation costs).

A careful reader will notice that a uniform backorder cost is assessed for all market segments. In reality, it is likely that different segments experience different backorder costs. However, it is unclear which segment should incur a higher backorder cost. For example, should Segment 1 have the highest backorder cost? To answer this question, notice that there are in fact two kinds of delays that are being considered here. One is *planned*, and the other *unplanned*. When a customer chooses to postpone the shipping date of his order, he accepts a planned delay. When the firm cannot ship his order on the date he has chosen, he experiences an unplanned delay. So, the question is whether a customer who cannot afford a planned delay (i.e., a bad planner) dislikes an unplanned delay more than a good planner or a patient consumer does? The answer is far from clear. Moreover, assigning different backorder costs to different market segments significantly complicates the analysis since the firm would then have to decide which order to ship when supply is insufficient. Consequently, I make the simplifying assumption that different market segments incur the same cost for *unplanned* delays.

3. Properties of an Optimal Price Schedule

Let $\{(p_k, \tau_k)\}_{k=0}^{K}$ for some nonnegative integer K be an optimal price schedule. Without loss of generality, we assume that if $K \geq 1$, then for each $k = 1, \ldots, K$, (p_k, τ_k) is chosen by at least one segment. Thus, $K \leq M$. Note that although $(p_0, \tau_0 = 0)$ is always available to the customers, it is possible that it does not draw any demand. A mechanism to break ties is also needed: If a customer derives the same surplus from two different options, he prefers the one with a lower price.

The following lemmas describe the characteristics of an optimal price schedule.

Lemma 1. *(i) $p_0 > p_1 > \ldots > p_K$. (ii) If segment m prefers (p_i, τ_i) to (p_j, τ_j) for some $i < j$, then segments $1, 2, \ldots, m - 1$ have the same preference. (iii) If $K \geq 1$, then there exist integers $\bar{1}, \ldots, \bar{K}$ with $1 \leq \bar{1} < \ldots < \bar{K} \leq M$ so that segments $1, \ldots, \bar{1} - 1$ choose (p_0, τ_0), segments $\bar{1}, \ldots, \bar{2} - 1$ choose (p_1, τ_1), etc., segments $\bar{K}, \ldots, M$ choose (p_K, τ_K).*

Proof. See Appendix A.

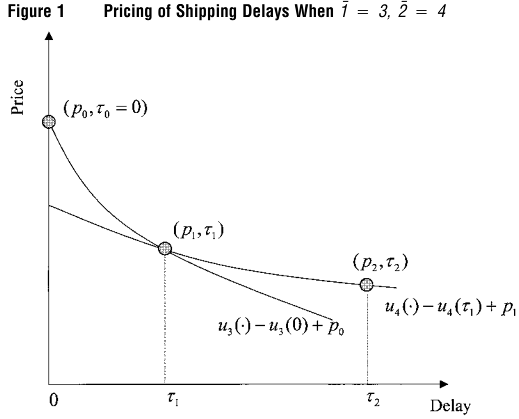
Lemma 1(i) is immediate: A longer delay should be priced lower. Because a lower segment (with a lower segment index) is more sensitive to waiting, it chooses a shorter shipping delay. This is the intuition behind (ii) and (iii).

Lemma 2. $u_{\bar{k}}(\tau_{k-1}) - p_{k-1} = u_{\bar{k}}(\tau_k) - p_k$, $k = 1, \ldots, K$.

Proof. See Appendix A.

For an illustration of Lemma 2, see Figure 1. The lemma indicates that the shipping delays should be priced so that the lowest segment for which a delivery option (i.e., price-delay combination) is meant, derives the same surplus from this "assigned" option as from the immediate lower option (i.e., shorter delay and

Figure 1 Pricing of Shipping Delays When $\bar{1} = 3$, $\bar{2} = 4$



higher price). For example, Segment $\bar{k}$ is the lowest segment for which (p_k, τ_k) is meant. From this option, Segment $\bar{k}$ gains surplus $u_{\bar{k}}(\tau_k) - p_k$, which is also the surplus it would gain from the immediate lower option (p_{k-1}, τ_{k-1}). Segment $\bar{k}$ chooses (p_k, τ_k) because of the tie-breaking mechanism. If we increase p_k, Segment $\bar{k}$ will switch to (p_{k-1}, τ_{k-1}); if we decrease it, the firm loses revenue. Therefore, the optimal price for a shipping delay is uniquely determined by the lowest segment that this delay is meant for and the price of the immediate shorter delay.

Lemma 2 also suggests that to solve the firm's optimization problem, we can restrict the decision variables to the shipping delays (τ_k's) and how they are assigned to the market segments ($\bar{k}$'s). These two decisions uniquely determine the prices.

4. Optimal Supply Chain Costs

Suppose the price schedule is given. Each customer self-selects a shipping delay and pays the corresponding price. Thus, the firm's long-run average revenues are fixed. Because no segments are priced out of the market, the long-run average variable costs are $\lambda_0 c$ per unit of time, which is also fixed. It only remains to minimize the firm's long-run average holding and backorder costs.

Under the given price schedule, let l_m be the shipping delay chosen by the customers from segment m, $m = 1, \ldots, M$. It is useful to distinguish between order and demand: If a type-m customer places an *order* at time t, the corresponding *demand* occurs at time $t + l_m$. Therefore, if $l_m > 0$ for some m, the firm has information about the future demand process. The question is how this information can be incorporated in the firm's replenishment policy. First, we need a notation to describe the demand information.

Take any segment $m = 1, \ldots, M$. Recall that orders from segment m arrive according to a Poisson process with average rate λ_m. The corresponding demand process is obtained by shifting the order process l_m units of time into the future. Let $D^m(t, t + w]$ be the total demand from segment m in the time interval $(t, t + w]$, $w \geq 0$. Thus, the firm at time t has already observed the value of $D^m(t, t + l_m]$. Define $x \wedge y = \min\{x, y\}$. Note that:

$$D^m(t, t + w] = D^m(t, t + w \wedge l_m] + D^m(t + w \wedge l_m, t + w],$$

where the first term on the right-hand side is realized by time t, but the second term remains uncertain at time t. In other words, the demand over any time interval can be decomposed to two parts, one realized at the start of the interval (hereafter, the *observable* part), and the other unknown at the start of the interval (hereafter, *unobservable* part). Clearly, $D^m(t, t + w \wedge l_m]$ and $D^m(t + w \wedge l_m, t + w]$ are independent Poisson random variables with means $\lambda_m(w \wedge l_m)$ and $\lambda_m(w - w \wedge l_m)$, respectively.

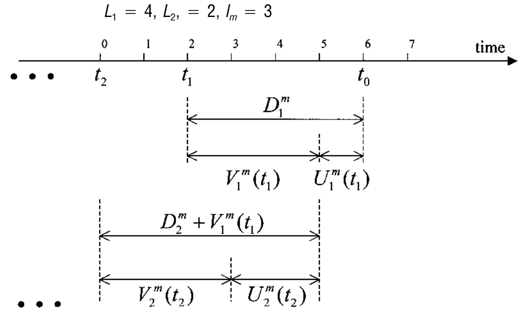
We now define a set of random variables that are useful in describing the demand uncertainties facing each stage of the supply chain. Take any Segment $m = 1, \ldots, M$. Let t_N be an arbitrary time epoch, and define recursively $t_n = t_{n+1} + L_{n+1}$ for $n = N - 1, \ldots, 1, 0$. Let D_n^m be the total demand from Segment m, in $(t_n, t_{n-1}]$, $n = 1, \ldots, N$. Define:

$$V_1^m(t_1) = D^m(t_1, t_1 + L_1 \wedge l_m], \text{ and}$$

$$U_1^m(t_1) = D^m(t_1 + L_1 \wedge l_m, t_0].$$

Thus, $V_1^m(t_1)$ and $U_1^m(t_1)$ are, respectively, the observable and unobservable parts of D_1^m (at time t_1). Note that $D_2^m + V_1^m(t_1) = D^m(t_2, t_1 + L_1 \wedge l_m]$, which can again be decomposed to two parts; let $V_2^m(t_2)$ and $U_2^m(t_2)$ be, respectively, the observable and unobservable parts. In general, let $V_n^m(t_n)$ and $U_n^m(t_n)$ be, respectively, the observable and unobservable parts of $D_n^m + V_{n-1}^m(t_{n-1})$, $n = 2, \ldots, N$. Figure 2 illustrates the definition of these random variables.

Figure 2 **Observable and Unobservable Random Variables When** $L_1 = 4$, $L_2 = 2$, $l_m = 3$



Define $\hat{L}_n = \Sigma_{i=1}^{n} L_i$, the total (downstream) leadtime at stage $n = 1, \ldots, N$, with $\hat{L}_0 = 0$. Recall that $V_1^m(t_1)$ is the observable part of $D^m(t_1, t_0]$. Also, $V_2^m(t_2)$ is the observable part of $D_2^m + V_1^m(t_1) = D^m(t_2, t_1 + L_1 \wedge l_m]$, which can be written as $D^m(t_2, t_2 + L_2 + L_1 \wedge l_m]$ since $t_1 = t_2 + L_2$. Thus,

$$V_2^m(t_2) = D^m(t_2, t_2 + (L_2 + L_1 \wedge l_m) \wedge l_m].$$

Because $(L_2 + L_1 \wedge l_m) \wedge l_m = \hat{L}_2 \wedge l_m$, $V_2^m(t_2)$ is the observable part of $D^m(t_2, t_2 + \hat{L}_2] = D^m(t_2, t_0]$. In general, one can show that $V_n^m(t_n)$ is the observable part of $D^m(t_n, t_0]$. Thus, the mean of $V_n^m(t_n)$ is $\lambda_m(\hat{L}_n \wedge l_m)$. Consequently, the mean of $D_n^m + V_{n-1}^m(t_{n-1})$ is $\lambda_m L_n + \lambda_m(\hat{L}_{n-1} \wedge l_m)$. Because $D_n^m + V_{n-1}^m(t_{n-1}) = V_n^m(t_n) + U_n^m(t_n)$ by definition, the mean of $U_n^m(t_n)$ is $\lambda_m(L_n + \hat{L}_{n-1} \wedge l_m - \hat{L}_n \wedge l_m)$. Finally, since $V_n^m(t_n)$ and $U_n^m(t_n)$ are by definition demands in nonoverlapping intervals, they are independent Poisson random variables.

Summing across the market segments, we define for any time epoch t, $U_n(t) = \Sigma_{m=1}^{M} U_n^m(t)$ and $V_n(t) = \Sigma_{m=1}^{M} V_n^m(t)$, $n = 1, \ldots, N$. From the above discussions and the fact that the demand processes from different market segments are independent, we have the following lemma.

Lemma 3. *For any time epoch t and any $n = 1, \ldots, N$, $U_n(t)$ and $V_n(t)$ are independent, $U_n(t)$ is Poisson with mean $\Sigma_{m=1}^{M} \lambda_m(\mathrm{L}_n + \mathrm{l}_m \wedge \hat{\mathrm{L}}_{n-1} - \mathrm{l}_m \wedge \hat{\mathrm{L}}_n)$, and $V_n(t)$ is Poisson with mean $\Sigma_{m=}^{M} \lambda_m (\mathrm{l}_m \wedge \hat{\mathrm{L}}_n)$. Moreover, at time t, the value of $V_n(t)$ is known, whereas the value of $U_n(t)$ remains uncertain.*

The optimal replenishment policy (that minimizes the long-run average holding and backorder costs in the supply chain) can be established by following Chen and Zheng (1994). The idea is to first develop a lower bound on the long-run average cost of any feasible policy, and then construct a feasible policy that achieves the lower bound. For the sake of brevity, the details are omitted. (A proof can be obtained from the author upon request.) Below, I briefly describe the optimal policy, the minimum systemwide cost, and some intuition for these results.

First, consider the following fictitious supply chain. There are N stages, and inventories flow from an outside supplier to Stage N, then to Stage $N - 1$, etc., and finally to Stage 1. The same holding and backorder costs are assessed as in the original supply chain. However, there is no longer advanced demand information, and the leadtime demand at Stage n is U_n, $n = 1, \ldots, N$. (U_n is a random variable identically distributed as $U_n(t)$ for any t.) This is the Clark-Scarf (1960) model, and the optimal policy is known to be of the echelon, base-stock type. The optimal base-stock levels are obtained by minimizing a sequence of convex functions. Define

$$G_1(y) = E[h_1(y - U_1) + (b + H_1)(y - U_1)^-].$$

This is a convex function. Let it be minimized at $y = y_1^*$. Now suppose $G_n(\cdot)$ is defined convex with a finite minimum point y_n^*. (True for $n = 1$.) Define

$$G_{n+1}(y) = E[h_{n+1}(y - U_{n+1}) + G_n(\min\{y_n^*, y - U_{n+1}\})],\ n = 1, \ldots, N - 1.$$

It is easily verified that $G_{n+1}(\cdot)$ is also convex with a finite minimum point y_{n+1}^*. The optimal base-stock level at stage n in the fictitious supply chain is y_n^*, $n = 1, \ldots, N$.

The above fictitious supply chain captures the demand uncertainties facing the original supply chain. In the original supply chain, Stage 1's leadtime demand is $D(t_1, t_0]$, the total demand in the interval, of which $V_1(t_1)$ is observed at time t_1, whereas $U_1(t_1)$ is not. The fictitious supply chain retains the unobserved part at Stage 1. The observed part $V_1(t_1)$ is shifted to Stage 2, and the "effective leadtime demand" at Stage 2 becomes $D(t_2, t_1] + V_1(t_1)$.[3] The unobserved part of this effective leadtime demand, $U_2(t_2)$ is again captured in the fictitious supply chain.

To describe the optimal policy, define the *echelon inventory position* at stage n to be the inventories in transit to or on hand at Stage n and all the downstream stages,

[3]It is quite intuitive why this is the effective leadtime demand at Stage 2. To see this, define the echelon inventory level at a stage to be its echelon inventory position minus the inventories in transit to the stage. Let $IL_n(t)$ be the echelon inventory level at Stage n at Time t. Consider Stage 2. The balance equation $IP_2(t_2) - D(t_2, t_1] = IL_2(t_1)$ is well known. Note that of $IL_2(t_1)$, $V_1(t_1)$ represents the committed inventory at Time t_1 (with customer names on them). Thus, the uncommitted or "safety" echelon inventory level at Stage 2 at Time t_1 is actually $IP_2(t_2) - D(t_2, t_1] - V_1(t_1)$. The difference between the echelon inventory position and the "safety" echelon inventory level is the effective leadtime demand.

minus the backorders at Stage 1. (A backorder at Stage 1 is an unsatisfied demand or, equivalently, a customer order that is not shipped on time.) Let $IP_n(t)$ be the echelon inventory position at Stage n at Time t.

The optimal policy for the original supply chain is obtained by combining the base-stock levels for the fictitious supply chain with the advanced demand information. Take any time epoch t. Let $v_n(t)$ be the realized value of $V_n(t)$, $n = 1, \ldots, N$. Define $s_n(t) = y_n^* + v_n(t)$, $n = 1, \ldots, N$. The following policy is optimal: For $n = 1, 2, \ldots, N$, whenever $IP_n(t)$ falls under $s_n(t)$, Stage $n + 1$ sends a shipment to Stage n to increase its echelon inventory position to $s_n(t)$, and if Stage $n + 1$ does not have sufficient on-hand inventory to achieve this goal, ship as much as possible. Because $v_n(t)$ changes over time, the policy is in fact a floating base-stock policy.

Theorem 1. *For any given price schedule, the optimal replenishment policy is to follow an echelon base-stock policy with order-up-to Level $s_n(t)$ at Time t for Stage n, $n = 1, \ldots, N$. The minimum long-run average systemwide holding and backorder costs are $C^* \overset{def}{=} G_N(y_N^*) + \sum_{n=1}^{N-1} h_{n+1} \sum_{m=1}^{M} \lambda_m(l_m \wedge \hat{L}_n)$.*[4]

As mentioned earlier, by choosing a shipping delay the customers provide the firm with advanced demand information. The value of this information in reducing the supply chain costs can be readily seen. For example, suppose $l_m \geq \hat{L}_N$ for all segments m. It can be easily verified that, in this case, $U_n \equiv 0$ for $n = 1, \ldots, N$, and thus, $G_N(y_N^*) = 0$. Therefore, $C^* = \sum_{n=1}^{N-1} h_{n+1}\lambda_0\hat{L}_n$, which represents the holding costs incurred for the inventories in transit. It is impossible to decrease the supply chain costs further without cutting the production/transportation leadtimes. In this example, there is no longer any uncertainty in the demand process as far as replenishment is concerned. In general, the longer the shipping delays (within the range $[0, \hat{L}_N]$), the lower the supply chain costs. However, the shipping delays are not without costs: Prices have to be discounted to entice customers to choose a shipping delay. Hence, the problem facing the firm is to balance between the lost revenues and the reduced supply chain costs.

[4] Hariharan and Zipkin (1995) have established this result for the case with a single market segment. I have generalized their results to multiple market segments with different shipping delays, which they refer to as demand leadtimes.

5. Computing an Optimal Price Schedule

Let $\{(p_k, \tau_k)\}_{k=0}^{K}$, $K \geq 0$ integer, be a price schedule with all the properties identified in §3. If $K = 0$, all customers pay the list price (p_0) with zero shipping delay ($\tau_0 = 0$). Otherwise, the market is segmented with Segments $1, \ldots, \bar{1} - 1$ choosing (p_0, τ_0), Segments $\bar{1}, \ldots, \bar{2} - 1$ choosing (p_1, τ_1), etc., Segments $\bar{K}, \ldots, M$ choosing (p_K, τ_K), where $\bar{1}, \bar{2}, \ldots, \bar{K}$ are integers with $1 \leq \bar{1} < \bar{2} < \ldots < \bar{K} \leq M$. Recall that the customers from Segment m arrive according to a Poisson process with average rate λ_m, $m = 1, \ldots, M$, and the arrival processes from different segments are independent. Therefore, the demands for each price-delay combination also arrive according to a Poisson process. Let λ^k be the average demand rate for (p_k, τ_k), $k = 0, 1, \ldots, K$. Define $\Lambda_i^j = \sum_{m=i}^{j} \lambda_m$ for any $i \leq j$. Thus, if $K = 0$, $\lambda^0 = \Lambda_1^M (= \lambda_0)$; if $K \geq 1$, $\lambda^0 = \Lambda_1^{\bar{1}-1}$, $\lambda^1 = \Lambda_{\bar{1}}^{\bar{2}-1}$, and so on.

Below, I develop several heuristic algorithms for computing an optimal price schedule. I begin with the special case where the supply chain has a single stage.

5.1. The Single-Stage Problem

Suppose $N = 1$. From Theorem 1, the minimum holding and backorder costs are determined by the distribution of U_1, a Poisson random variable with mean $\sum_{k=0}^{K} \lambda^k (L - \tau_k \wedge L)$. (For convenience, we write L for L_1 in this section.) Because a shipping delay larger than L does not provide any additional reduction in the holding and backorder costs, we restrict to $\tau_k \leq L$ for all k. As a result,

$$E[U_1] = \sum_{k=0}^{K} \lambda^k (L - \tau_k) = \lambda_0 L - \sum_{k=1}^{K} \lambda^k \tau_k.$$

Let $C(\sum_{k=1}^{K} \lambda^k \tau_k)$ be the minimum holding and backorder costs.

We can formulate the problem of finding an optimal price schedule as follows:[5]

[5] As mentioned earlier, the long-run average variable costs are equal to $c\lambda_0$, which is independent of decisions.

$$\begin{aligned}\max_{0\le\tau_1\le\ldots\le\tau_K\le L} \ & \sum_{k=0}^{K} p_k\lambda^k - C\Big(\sum_{k=1}^{K}\lambda^k\tau_k\Big)\\ \text{s.t. } & u_{\bar 1}(0) - p_0 = u_{\bar 1}(\tau_1) - p_1\\ & u_{\bar 2}(\tau_1) - p_1 = u_{\bar 2}(\tau_2) - p_2\\ & \vdots\\ & u_{\bar K}(\tau_{K-1}) - p_{K-1} = u_{\bar K}(\tau_K) - p_K,\end{aligned} \tag{3}$$

where the constraints are from Lemma 2. The decision variables are K (≥ 1), $(\bar 1, \ldots, \bar K)$, which are integers with $1 \le \bar 1 < \bar 2 < \ldots < \bar K \le M$, $(p_1, \ldots, p_K)$, and $(\tau_1, \ldots, \tau_K)$. Note that if $K = 1$ and $\tau_1 = 0$, then only $(p_0, 0)$ is offered in the price schedule. Therefore, the value of K does not necessarily represent the number of positive shipping delays. This use of K is slightly different from its previous usage.

Lemma 4. *If $0 < \tau_{K-1} < \tau_K < L$, then the price schedule is not optimal.*

Proof. See Appendix A.

Therefore, if an optimal price schedule offers two or more positive shipping delays, then the largest shipping delay must be L. This suggests a two-step algorithm.

Algorithm 1-I:

Suppose at most one positive shipping delay is offered. In this case, we can restrict to $K = 1$. For convenience, replace (p_1, τ_1) with (p, τ) and $\bar 1$ with i. Then the optimization problem in (3) reduces to:

$$\begin{aligned}\max \ & \Lambda_1^{i-1}p_0 + \Lambda_i^M p - C(\Lambda_i^M\tau)\\ \text{s.t. } & u_i(0) - p_0 = u_i(\tau) - p,\end{aligned}$$

where the decision variables are $i = 1, \ldots, M$, τ with $0 \le \tau \le L$, and p. Note that if $\tau = 0$, then only $(p_0, 0)$ is offered. Using the constraint to express p as a function of τ and i, rewrite the above problem as:

$$\max_{i,\tau} \Lambda_1^{i-1}p_0 + \Lambda_i^M(u_i(\tau) - u_i(0) + p_0) - C(\Lambda_i^M\tau).$$

This is an easy problem.

Now suppose at least two positive shipping delays are offered. From Lemma 4, the shipping delay "assigned" to Segment M must be L. I suggest the following procedure to determine a heuristic solution to (3). Suppose Segments $1, \ldots, M-1$ are initially assigned a shipping delay of 0. In iteration 1, only Segment $M - 1$ is allowed to "leave" its initial assignment, whereas the shipping delays assigned to all the other segments remain fixed. This is a one-dimensional problem (with only one decision variable). Let τ be the optimal shipping delay for Segment $M - 1$ (that maximizes the firm's profits). If $\tau = 0$, i.e., Segment $M - 1$ stays with its initial assignment, then the procedure stops and the initial assignment is the solution. Otherwise, if $\tau > 0$, then proceed to Iteration 2, where Segment $M - 2$ is allowed to leave its initial assignment. In this iteration, the shipping delays assigned to all the other segments remain fixed, and the shipping delay for Segment $M - 2$ is confined to $[0, \tau]$. The algorithm stops if Segment $M - 2$ stays at 0 and continues otherwise. Due to limited space here, I omit the details, which can be found in Chen (1999).

An optimal price schedule is much easier to find if the reservation-price functions and the supply-chain cost function have some additional properties. I now present these properties and the resulting algorithm.

Theorem 2. *If $u_m(\tau) - u_{m+1}(\tau)$ is convex in τ for $m = 1, \ldots, M-1$ and $C(\cdot)$ is concave, then it is not optimal to offer a shipping delay strictly between 0 and L.*

Proof. From the constraints in (3), we have for any $k = 1, \ldots, K$,

$$\begin{aligned}p_k &= \sum_{j=1}^{k}(u_{\bar j}(\tau_j) - u_{\bar j}(\tau_{j-1})) + p_0\\ &= u_{\bar k}(\tau_k) + \sum_{j=1}^{k-1}(u_{\bar j}(\tau_j) - u_{\overline{j+1}}(\tau_j)) - u_{\bar 1}(0) + p_0.\end{aligned}$$

Because $u_m(\cdot) - u_{m+1}(\cdot)$ is convex for $m = 1, \ldots, M-1$, $u_i(\cdot) - u_j(\cdot)$ is convex for any $i < j$. Therefore, the revenue part of the objective function in (3) is convex in $(\tau_1, \ldots, \tau_K)$. Because $C(\cdot)$ is concave, $C(\Sigma_{k=1}^K \lambda^k\tau_k)$ is concave in $(\tau_1, \ldots, \tau_K)$. Consequently, the objective function in (3) is convex in $(\tau_1, \ldots, \tau_K)$. Consequently, the optimal shipping delays can only be at the end points of the interval $[0, L]$. □

Note that if the reservation-price functions are given by (1) or (2), then $u_m(\cdot) - u_{m+1}(\cdot)$ is convex for all m. On the other hand, the minimum cost function $C(\cdot)$ is concave under a well-known approximation. In the standard single-stage inventory problem with a base-stock policy, a good approximation of the minimum holding and backorder costs is a linear, increasing function of the standard deviation of the leadtime demand (e.g., Zipkin 2000). (This is, in fact, exact if the leadtime demand is a normal random variable.) Thus, $C(\Sigma_{k=1}^{K}\lambda^k\tau_k) \approx \alpha\sqrt{\lambda_0 L - \Sigma_{k=1}^{K}\lambda^k\tau_k}$ for some positive constant α, since $\lambda_0 L - \Sigma_{k=1}^{K}\lambda^k\tau_k$, as mentioned earlier, is the mean (and thus the variance) of the Poisson random variable U_1, which serves as the "leadtime demand." Under this approximation, $C(\Sigma_{k=1}^{K}\lambda^k\tau_k)$ is concave in $(\tau_1, \ldots, \tau_K)$.

Algorithm 1-II:

Suppose we restrict shipping delays to 0 and L. In this case, the optimal price schedule can be characterized by a single variable i that serves to partition the market: Segments $1, \ldots, i-1$ choose 0 shipping delay, and the remaining segments choose L, $i = 1, \ldots, M+1$. If $i = 1$, all segments choose L; if $i = M+1$, all segments choose 0. The price paid by the segments choosing L is $p = p_0 - u_i(0) + u_i(L)$ (Lemma 2). The following is the firm's profit:

$$p_0\Lambda_1^{i-1} + p\Lambda_i^M - C(\Lambda_i^M L).$$

Maximizing the above expression over i yields the solution.

5.2. The Multistage Problem

Suppose $N \geq 2$, i.e., the supply chain consists of multiple stages. From § 4, the minimum systemwide holding and backorder costs are determined by the distributions of $U_n(t)$ and $V_n(t)$, $n = 1, \ldots, N$. From Lemma 3, these are Poisson random variables with

$$\mu_n^u \stackrel{def}{=} E[U_n(t)] = \sum_{k=0}^{K} \lambda^k(L_n + \tau_k \wedge \hat{L}_{n-1} - \tau_k \wedge \hat{L}_n) \tag{4}$$

and

$$\mu_n^v \stackrel{def}{=} E[V_n(t)] = \sum_{k=0}^{K} \lambda^k(\tau_k \wedge \hat{L}_n). \tag{5}$$

(The above expressions are based on the fact that $l_m = \tau_k$ for $m = \bar{k}, \ldots, \overline{k+1} - 1$ with $\bar{0} = 1$.) Because $G_N(y_N^*)$ is a function of $(\mu_1^u, \ldots, \mu_N^u)$, we write $C(\mu_1^u, \ldots, \mu_N^u)$ for $G_N(y_N^*)$. Thus,

$$C^* = C(\mu_1^u, \ldots, \mu_N^u) + \sum_{n=1}^{N-1} h_{n+1}\mu_n^v. \tag{6}$$

It can be easily verified that increasing a shipping delay beyond $\hat{L}_N$ no longer affects μ_n^u and μ_n^v for any n and thus does not further decrease the supply chain costs. Therefore, we restrict to $\tau_k \leq \hat{L}_N$ for all k. As in (3), the problem of finding an optimal price schedule can be stated as:

$$\begin{aligned}
\max_{0 \leq \tau_1 \leq \ldots \leq \tau_K \leq \hat{L}_N} \; & \sum_{k=0}^{K} p_k\lambda^k - C(\mu_1^u, \ldots, \mu_N^u) - \sum_{n=1}^{N-1} h_{n+1}\mu_n^v \\
\text{s.t. } & u_{\bar{1}}(0) - p_0 = u_{\bar{1}}(\tau_1) - p_1 \\
& u_{\bar{2}}(\tau_1) - p_1 = u_{\bar{2}}(\tau_2) - p_2 \\
& \vdots \\
& u_{\bar{K}}(\tau_{K-1}) - p_{K-1} = u_{\bar{K}}(\tau_K) - p_K,
\end{aligned} \tag{7}$$

where the decision variables are $K \geq 1$, $(\bar{1}, \ldots, \bar{K})$ with $1 \leq \bar{1} < \ldots < \bar{K} \leq M$, $(p_1, \ldots, p_K)$, and $(\tau_1, \ldots, \tau_K)$. Below, I develop two algorithms for solving this problem; they parallel Algorithms 1-I and 1-II for the single-stage problem.

With proper restrictions, the optimization problem (7) reduces to a single-stage problem. For example, this is true if all shipping delays are confined to $[0, L_1]$. For a more sophisticated example, take any $i = 1, \ldots, M$ and any $n \geq 2$. Suppose we fix the choices by Segments $1, \ldots, i-1$. Assume $l_m \leq \hat{L}_{n-1}$ for $m = 1, \ldots, i-1$. Under the restriction that Segments $i, \ldots, M$ can only choose shipping delays in $[\hat{L}_{n-1}, \hat{L}_n]$, the firm's (local) optimization problem again reduces to a single-stage problem (see Chen 1999 for details). The following lemma is an immediate result of this observation:

LEMMA 5. *If $\hat{L}_{n-1} < \tau_{K-1} < \tau_K < \hat{L}_n$ for some $n = 1, \ldots, N$, then the price schedule is not optimal.*

PROOF. Take any feasible solution to (7) with $\hat{L}_{n-1} < \tau_{K-1} < \tau_K < \hat{L}_n$ for some $n = 1, \ldots, N$. Consider the local optimization problem for the interval $[\hat{L}_{n-1}, \hat{L}_n]$. The proof is essentially the same as that of Lemma 4. □

Algorithm N-I:

First, solve (7) with the additional constraint that all shipping delays are in $[0, \hat{L}_1]$. This is a single-stage problem and can thus be solved by using Algorithm 1-I. In the solution, if Segments $m, \ldots, M$ for some $m \leq M$ are assigned a shipping delay equal to $\hat{L}_1$, then the algorithm continues to the next step, otherwise the algorithm stops and the resulting price schedule is the solution. Suppose the algorithm continues. Fix the choices (i.e., the shipping delays and the corresponding prices) of Segments $1, \ldots, m-1$, and allow Segments $m, \ldots, M$ to increase their shipping delays from the current value $\hat{L}_1$ up to $\hat{L}_2$. This local optimization problem can again be solved by using Algorithm 1-I. If in the solution some segments choose a shipping delay equal to $\hat{L}_2$, then continue and allow these segments to increase their shipping delays from $\hat{L}_2$ up to $\hat{L}_3$, otherwise stop, and so on.

The following theorem is parallel to Theorem 2 for the single-stage problem.

THEOREM 3. *If $u_m(\tau) - u_{m+1}(\tau)$ is convex in τ for all m, and $C(\mu_1^u, \ldots, \mu_N^u)$ is concave and twice-differentiable in μ_n^u, $n = 1, \ldots, N$, then there exists an optimal price schedule that only provides shipping delays from the set $\{\hat{L}_n, n = 0, 1, \ldots, N\}$.*

PROOF. Suppose $u_m(\cdot) - u_{m+1}(\cdot)$ is convex for all m, and $C(\mu_1^u, \ldots, \mu_N^u)$ is concave and twice-differentiable in each of its arguments. As in the proof of Theorem 2, one can express the revenue term $\Sigma_{k=0}^K p_k \lambda^k$ as a function of $(\tau_1, \ldots, \tau_K)$ based on the constraints, and show that this function is convex in τ_k, $k = 1, \ldots, K$. Now consider the cost terms in (7). Take any $k = 1, \ldots, K$. From (5), μ_n^v is concave in τ_k for all n. Below, we show that the cost function $C(\mu_1^u, \ldots, \mu_N^u)$ is concave in τ_k for $\tau_k \in [\hat{L}_{n-1}, \hat{L}_n]$, $n = 1, \ldots, N$. First, note that for any $\tau_k \in [\hat{L}_{n-1}, \hat{L}_n]$,

$$\frac{\partial \mu_i^u}{\partial \tau_k} = 0, \ \forall i \neq n; \text{ and } \frac{\partial \mu_n^u}{\partial \tau_k} = -\lambda^k.$$

Therefore for any $\tau_k \in [\hat{L}_{n-1}, \hat{L}_n]$,

$$\frac{\partial C}{\partial \tau_k} = \sum_{i \neq n} \frac{\partial C}{\partial \mu_i^u} \frac{\partial \mu_i^u}{\partial \tau_k} + \frac{\partial C}{\partial \mu_n^u} \frac{\partial \mu_n^u}{\partial \tau_k} = -\lambda^k \frac{\partial C}{\partial \mu_n^u}$$

and

$$\frac{\partial^2 C}{\partial \tau_k^2} = -\lambda^k \frac{\partial^2 C}{\partial (\mu_n^u)^2} \frac{\partial \mu_n^u}{\partial \tau_k} = (\lambda^k)^2 \frac{\partial^2 C}{\partial (\mu_n^u)^2} \leq 0$$

since C is concave in μ_n^u. Consequently, the objective function in (7) is convex in τ_k for $\tau_k \in [\hat{L}_{n-1}, \hat{L}_n]$, $k = 1, \ldots, K$ and $n = 1, \ldots, N$, and, thus, the optimal shipping delays can only be at the end points of these intervals. □

There is some numerical evidence that the minimum cost for serial inventory systems is concave and increasing in the variance of the leadtime demand at each stage (see Chen 1998). In this setting, since the demand process is Poisson, μ_n^u is the mean (thus variance) of the leadtime demand at Stage n. This supports the assumption that $C(\mu_1^u, \ldots, \mu_N^u)$ is concave in μ_n^u, $n = 1, \ldots, N$.

Algorithm N-II:
Suppose we restrict the shipping delays to the set of all total leadtimes, $\{\hat{L}_n\}_{n=0}^N$. In this case, we can characterize an optimal solution by a set of partition integers, δ_n for $n = 1, \ldots, N$, whereby Segments $\delta_n, \ldots, M$ choose a shipping delay greater than or equal to $\hat{L}_n$. Therefore, Segments $1, \ldots, \delta_1 - 1$ choose zero shipping delay, Segments $\delta_n, \ldots, \delta_{n+1} - 1$ choose $\hat{L}_n$ for $n = 1, \ldots, N-1$, and Segments $\delta_N, \ldots, M$ choose $\hat{L}_N$. The partition integers satisfy $1 \leq \delta_1 \leq \delta_2 \leq \ldots \leq \delta_N \leq M + 1$. If $\delta_n = \delta_{n+1}$ for some n, then no segments choose $\hat{L}_n$; if $\delta_n = M + 1$, then no segments choose shipping delays larger than or equal to $\hat{L}_n$. Therefore, the partition integers uniquely determine the shipping delays offered, how these delays are assigned to market segments, and by Lemma 2, their corresponding prices. It only remains to determine the partition integers.

For small values of M and N, it may be feasible to search over all possible combinations of the partition integers for an optimal solution. If this is impossible, we suggest a simple sequential procedure. To determine δ_1, restrict the shipping delays to either 0 ($= \hat{L}_0$) or $\hat{L}_1$. This is a simple, one-dimensional problem. If the solution is $\delta_1 = M + 1$, then the procedure ends. Otherwise, proceed to determine δ_2. This is achieved by assuming that Segments $\delta_1, \ldots, M$ can now choose between $\hat{L}_1$ and $\hat{L}_2$ while fixing the choices by Segments $1, \ldots, \delta_1 - 1$ (they all choose 0). This is again an easy problem. If the solution is $\delta_2 = M + 1$, then stop; otherwise, continue to the next step, and so on.

6. Numerical Examples

The premise of this paper is that customers exhibit different degrees of aversion to shipping delays. This provides an opportunity for the firm to segment the market so that different customers choose different shipping delays and pay different prices. The delays in shipping customer orders allow the firm to reduce its inventories. This section uses numerical examples to illustrate the benefits of this idea. As in any numerical studies, the results are suggestive, not conclusive.

I now describe the numerical examples. Suppose the reservation-price functions are as given in (1). The numerical examples have the following parameters:

$$p_0 = 10.$$
$$M = 2.$$
$$u_m(0) = 15 \text{ for } m = 1,2.$$
$$(\lambda_1, \lambda_2) = (1,9),(5,5),(9,1).$$
$$(\theta_1, \theta_2) = (0.2, 0.1),(1, 0.1),(1, 0.9).$$
$$N = 1,2,3.$$
$$L_n = L/N \text{ for } n = 1,\ldots,N, \text{ where } L = 1,3,5.$$
$$h_n = 1/N \text{ for } n = 1,\ldots,N.$$
$$b = 5,10,15.$$

There are 243 examples in total. (As mentioned earlier, the firm's long-run average variable costs are fixed. I, thus, ignore them in the numerical study.)

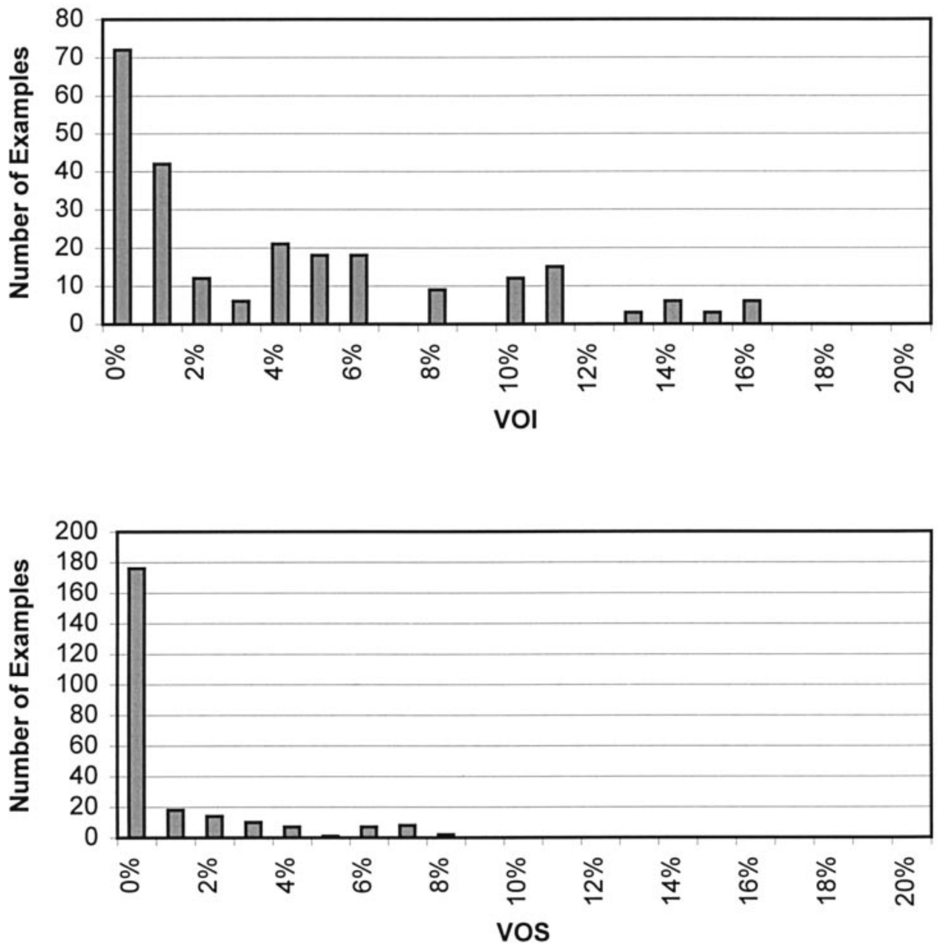
For each example, I considered the following three scenarios: (1) The status quo is where only (p_0, 0) is offered, i.e., the firm offers only the list price and promises all customers that their orders will be shipped immediately. (2) Only one option is available to all segments, i.e., all customers pay the same price p and choose the same shipping delay τ, which may be positive. Here the firm's decisions are (p, τ), given that the customers have already seen the list price. (Again, no segments are priced out of the market in the second stage of selling.) Therefore, in this scenario, the firm recognizes that customers can be enticed to accept a shipping delay, but ignores the fact that there are different market segments with different degrees of aversion to shipping delays. (3) The firm offers a price schedule, $\{(p_k,\ \tau_k)\}_{k=0}^{K}$, and the customers self-select which price to pay and which shipping delay to take. For this problem, I used Algorithm N-II to determine the price schedule. Let the firm's long-run average profits be π_i, under Scenario i, i = 1, 2, 3. It is clear that $\pi_1 \le \pi_2 \le \pi_3$. The relative increase from π_1 to π_2 represents the value of information (VOI) since the firm benefits by extracting advanced demand information, and the relative increase from π_2 to π_3 represents the value of segmentation (VOS) since the firm recognizes the potential of offering different options to different market segments. That is,

$$\text{VOI} = \frac{\pi_2 - \pi_1}{\pi_1} \text{ and } \text{VOS} = \frac{\pi_3 - \pi_2}{\pi_2}.$$

The results are summarized in Figures 3 and 4. Figure 3 contains the histograms for VOI and VOS across all examples. The average VOI is 3.73%, and the average VOS is 0.75%. It appears that advanced demand information is very valuable, whereas segmenting the market to exploit its microstructure provides a relatively marginal value. (Note that if we include the firm's variable costs, the profits under the three scenarios will decrease by the same amount, increasing both VOI and VOS. Therefore, a firm with a thin profit margin stands to profit significantly from the above strategies.) Figure 4 shows how VOI and VOS change as various system parameters are varied, suggesting the following:

- VOI depends only on the total market size, whereas VOS depends on the relative sizes of the market segments. This is not surprising: The firm views the market as a single entity in Scenario 2, but it treats the segments differently in Scenario 3, and the benefit of segmentation is larger if there are more customers in the higher segment (more patient).
- VOI is higher when customers are more patient, and VOS is higher when segments are more different in terms of their aversion to shipping delays. This is again not surprising.
- VOI decreases, and VOS increases, as N increases. Here is an intuitive explanation of this phenomenon. As we increase the number of stages in the supply chain (while fixing the total leadtime from the outside supplier to Stage 1, the last stocking point), there are more intermediate stocking points. And it is less costly to hold inventories at these intermediate points than at Stage 1. The consequence is that the supply chain is better positioned to cope with demand uncertainty. Therefore, as N increases, the incremental value of information diminishes, i.e., VOI decreases. On the other hand, when N is small, say $N = 1$, there is little flexibility to treat the different segments differently. (By Theorem 2, the shipping delay can only be 0 or L.) As N increases, this flexibility increases. So does VOS.
- As the (total) leadtime increases, VOI increases, and VOS exhibits an increasing trend.

Figure 3 **Histograms**

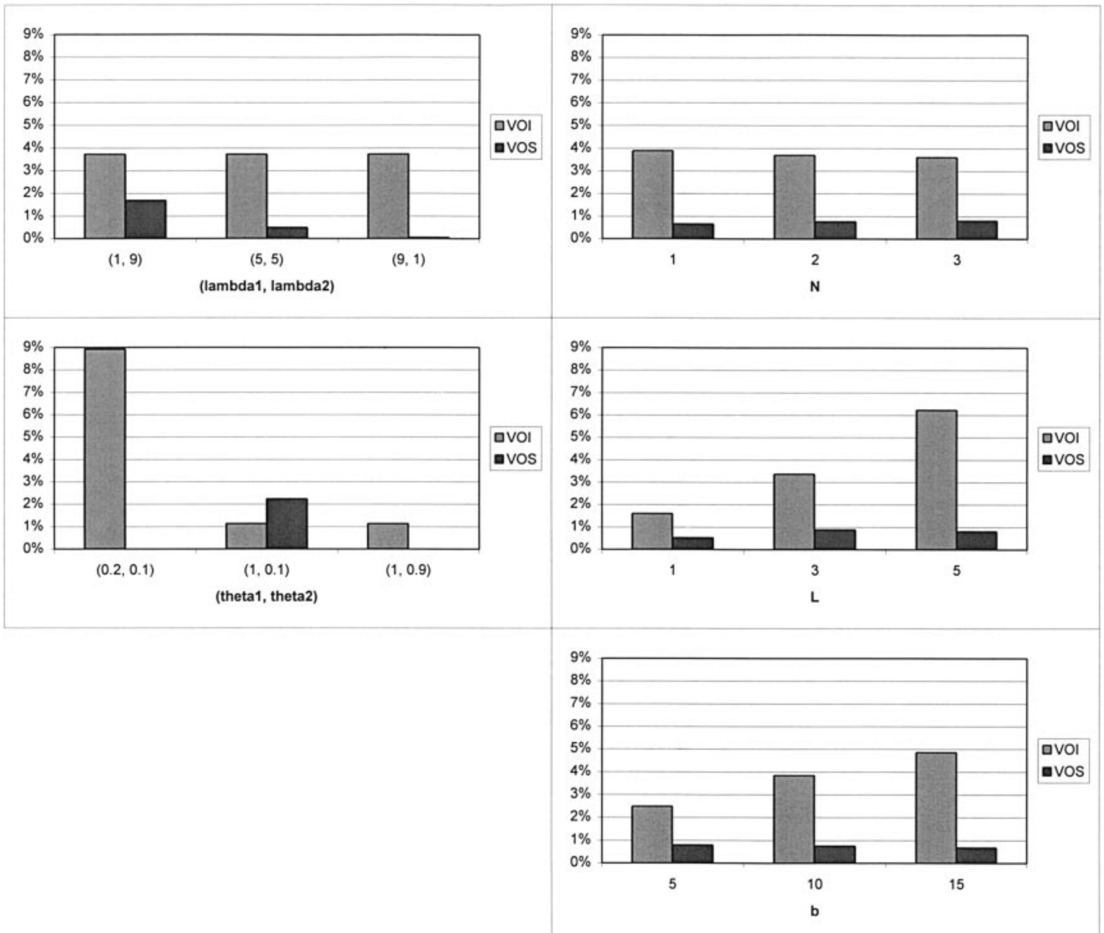


• As b (the backorder penalty cost) increases, VOI increases but VOS decreases. In other words, as customer service becomes more important (i.e., higher b), the value of advanced demand information increases. On the other hand, since the different market segments share the same backorder penalty cost (they are equally averse to unplanned delays), the segments become more alike as b increases, and thus, VOS decreases.

7. Conclusion

It is often true that some customers may be willing to accept a delay in shipping if they are given a price discount in return. I have studied a model where a monopolist sells a single product to a heterogeneous market with several segments exhibiting different degrees of aversion to shipping delays. The firm offers a menu of price-delay combinations, from which the

Figure 4 **Value of Information/Segmentation and System Parameters**



customers each select one that maximizes their surplus. When a customer chooses a positive shipping delay, i.e., a positive lag from when the order is placed to when it is to be shipped, the firm gains advanced demand information. This information can then be used to reduce inventories. The paper shows how the firm can determine an optimal price-delay schedule.

I have also provided numerical examples to illustrate the benefits of the above pricing strategy. For each numerical example, I considered the following three scenarios. (1) The firm offers only a list price and promises all customers that their orders will be shipped immediately. (2) The firm offers one price and one shipping delay for all market segments. The firm thus treats the market as a homogeneous body that can be enticed to accept a shipping delay. (3) The firm offers a menu of price-delay combinations, and different market segments can choose different options. I called the relative increase in the firm's profits from Scenario 1 to 2 the value of information, and from Scenario 2 to

3 the value of segmentation. Section 6 illustrates the magnitudes of these values and how they depend on several market and supply chain characteristics.

Several extensions of the current model are possible. The current assumption of simple Poisson processes can be replaced easily by compound Poisson processes without affecting the form of the optimal replenishment policy. Similarly, we can replace the serial supply chain structure with an assembly structure, and allow the inventory transfers from one stage to another to be integer multiples of a fixed, stage-specific batch size. (These extensions are rather straightforward by following Chen and Zheng 1994 and Chen 2000.) However, extensions to other supply chain structures such as distribution networks are less obvious, since it is unclear how the advanced demand information can be utilized optimally.

Acknowledgments

The author would like to thank the editors and referees for their comments and suggestions, which have significantly improved this paper. Financial support from the National Science Foundation, the Columbia Business School, and the Eugene Lang Foundation is gratefully acknowledged.

Appendix A: Omitted Proofs

Proof of Lemma 1. (i) Suppose, to the contrary, there exist i and j with $i < j$ and $p_i \leq p_j$. Because $\tau_i < \tau_j$, it is easy to verify that all customers prefer (p_i, τ_i) to (p_j, τ_j). The latter pair, thus, draws no demand, a contradiction.

(ii) Because Segment m prefers (p_i, τ_i) to (p_j, τ_j), its surplus with the former pair is higher, i.e.,

$$u_m(\tau_i) - p_i > u_m(\tau_j) - p_j. \tag{8}$$

Now take any Segment $m' < m$. Because $u_m(\cdot)$ is flatter than $u_{m'}(\cdot)$, $u_m(\tau_i) - u_m(\tau_j) < u_{m'}(\tau_i) - u_{m'}(\tau_j)$. Combining this inequality with (8), we see that Segment m' achieves a higher surplus with (p_i, τ_i) than with (p_j, τ_j).

(iii) It follows directly from (ii). □

Proof of Lemma 2. First, suppose, to the contrary, $u_{\bar{1}}(\tau_0) - p_0 \neq u_{\bar{1}}(\tau_1) - p_1$. Because Segment $\bar{1}$ by definition prefers (p_1, τ_1) to all the other combinations, it is impossible to have $u_{\bar{1}}(\tau_1) - p_1 < u_{\bar{1}}(\tau_0) - p_0$. Now suppose:

$$u_{\bar{1}}(\tau_0) - p_0 < u_{\bar{1}}(\tau_1) - p_1. \tag{9}$$

In this case, increase p_1 so that (9) becomes an equality. Let δ be the increment. Let $p'_k = p_k + \delta$ for $k = 1, 2, \ldots, K$. Consider the new price schedule $\{(p_0, \tau_0), (p'_k, \tau_k)_{k=1}^K\}$. Note that although Segment $\bar{1}$ derives the same surplus from (p'_1, τ_1) and (p_0, τ_0), it prefers the former due to the tie-breaking mechanism. Moreover, Segment $\bar{1}$ prefers (p'_1, τ_1) to (p'_k, τ_k) for any $k \geq 2$ since it preferred (p_1, τ_1) to (p_k, τ_k) in the old price schedule and $p'_1 - p_1 = p'_k - p_k$. Consequently, Segment $\bar{1}$ chooses (p'_1, τ_1) under the new price schedule. For any segment $m < \bar{1}$, since Segment m preferred (p_0, τ_0) to the other options in the old price schedule and those other options are now less attractive because of their higher prices, Segment m's choice remains the same under the new price schedule. On the other hand, for any segment $m > \bar{1}$, since $u_m(\cdot)$ is flatter than $u_{\bar{1}}(\cdot)$, Segment m prefers (p'_1, τ_1) to (p_0, τ_0), and therefore, can restrict its choice to the set $\{(p'_k, \tau_k)\}_{k=1}^K$. Because $p'_k - p_k$ is the same for all $k \geq 1$, Segment m chooses the same shipping delay as before. In sum, the new price schedule does not change the shipping delay chosen by any of the market segments and thus does not affect the firm's operational costs. But, it increases the revenues. Therefore, the old price schedule is not optimal, a contradiction. Thus, (9) must be an equality. The rest of the proof is similar. □

Proof of Lemma 4. Take any feasible solution to (3), $\{(p_k, \tau_k)\}_{k=1}^K$, with $0 < \tau_{K-1} < \tau_K < L$. (Thus, $K \geq 2$.) Now suppose we want to maximize the firm's profits by only changing τ_{K-1} and τ_K subject to an additional constraint that $(\lambda^{K-1}\tau_{K-1} + \lambda^K\tau_K)$ remains fixed. This additional constraint implies that there is really just one variable, say, τ_{K-1}, with $\partial\tau_K/\partial\tau_{K-1} = -\lambda^{K-1}/\lambda^K$, and that the firm's operational costs remain fixed. After ignoring the other parts in the objection function that do not depend on τ_{K-1}, we have the objection function as:

$$\psi \stackrel{def}{=} p_{K-1}\lambda^{K-1} + p_K\lambda^K,$$

where both p_{K-1} and p_K are functions of τ_{K-1}. These functions can be derived from the constraints in (3): $u_{\overline{K-1}}(\tau_{K-2}) - p_{K-2} = u_{\overline{K-1}}(\tau_{K-1}) - p_{K-1}$ and $u_{\bar{K}}(\tau_{K-1}) - p_{K-1} = u_{\bar{K}}(\tau_K) - p_K$. Substituting these functions for p_{K-1} and p_K, and ignoring the fixed parts, in the objective function, we have

$$\begin{aligned}\psi &= (\lambda^{K-1} + \lambda^K)u_{\overline{K-1}}(\tau_{K-1}) \\ &\quad - \lambda^K u_{\bar{K}}(\tau_{K-1}) + \lambda^K u_{\bar{K}}(\tau_K).\end{aligned}$$

Note that:

$$\begin{aligned}\frac{\partial\psi}{\partial\tau_{K-1}} &= (\lambda^{K-1} + \lambda^K)u'_{\overline{K-1}}(\tau_{K-1}) - \lambda^K u'_{\bar{K}}(\tau_{K-1}) \\ &\quad + \lambda^K u'_{\bar{K}}(\tau_K)\frac{\partial\tau_K}{\partial\tau_{K-1}} \\ &= (\lambda^{K-1} + \lambda^K)u'_{\overline{K-1}}(\tau_{K-1}) \\ &\quad - \lambda^K u'_{\bar{K}}(\tau_{K-1}) - \lambda^{K-1}u'_{\bar{K}}(\tau_K) \\ &< (\lambda^{K-1} + \lambda^K)u'_{\bar{K}}(\tau_{K-1}) \\ &\quad - \lambda^K u'_{\bar{K}}(\tau_{K-1}) - \lambda^{K-1}u'_{\bar{K}}(\tau_K) \\ &= \lambda^{K-1}[u'_{\bar{K}}(\tau_{K-1}) - u'_{\bar{K}}(\tau_K)] \\ &< 0,\end{aligned}$$

where the first inequality follows since both $u_{\bar{K}}(\cdot)$ and $u_{\overline{K-1}}(\cdot)$ are decreasing, with the former being flatter than the latter, and the second inequality follows since $u_{\bar{K}}(\cdot)$ is strictly convex and $\tau_{K-1} < \tau_K$. Therefore, the current solution is not optimal. □

References

Chen, F., Y.-S. Zheng. 1994. Lower bounds for multi-echelon stochastic inventory systems. *Management Sci.* **40** 1426–1443.

——. 2000. Optimal policies for multi-echelon inventory problems with batch ordering. *Oper. Res.* **48**(3) 376–389.

——. 1998. On (R,nQ) policies in serial inventory systems. S. Tayur, R. Ganeshan, M. Magazine, eds *Quantitative Models for Supply Chain Management. Kluwer Academic Publishers, the Netherlands.*

——. 1999. Market segmentation, advanced demand information, and supply chain performance. Working paper. Columbia University, New York, NY.

——, J. Eliashberg, P. Zipkin. 1998. Customer preferences, supply-chain costs, and product-line design. T.-H. Ho, C. S. Tang, eds. *Product Variety Management: Research Advances.* Kluwer Academic Publishers, the Netherlands.

Clark, A., H. Scarf. 1960. Optimal policies for a multi-echelon inventory problem. *Management Sci.* **6** 475–490.

de Groote, X. 1994. Flexibility and product variety in lot-sizing models. *Eur. J. Oper. Res.* **75** 264–274.

Eliashberg, J. A. Manrai. 1989. Optimal positioning of new product concepts: Some analytical implications and empirical results. *Eur. J. Oper. Res.* **63** 376–397.

Haley, R. 1968. Benefit segmentation: A decision-oriented research tool. *J. Marketing* **32** 30–35.

Hariharan, R., P. H. Zipkin. 1995. Customer-order information, leadtimes, and inventories. *Management Sci.* **41** 1599–1607.

Hauser, J., P. Simmie. 1981. Profit maximizing perceptual positions: An integrated theory for the selection of product features and price. *Management Sci.* **27** 33–56.

Ho, T.-H., C. S. Tang. 1998. *Product Variety Management: Research Advances.* Kluwer Academic Publishers, the Netherlands.

Hotelling, H. 1929. Stability in competition. *Econ. J.* **39** 41–57.

Lancaster, K. 1979. *Variety, Equity and Efficiency.* Columbia University Press, New York, NY.

——. 1990. The economics of product variety: A survey. *Marketing Sci.* **9**(3) 189–206.

Lovejoy, W., S. Whang. 1995. Response time design in integrated order processing/production systems. *Oper. Res.* **43** 851–861.

Mendelson, H., S. Whang. 1990. Optimal incentive-compatible priority pricing for the M/M/1 queue. *Oper. Res.* **38** 870–883.

Milgrom, P., J. Roberts. 1988. Communication and inventory as substitutes in organizing production. *Scandinavian J. Econ.* **90** 275–289.

Moorthy, K. 1984. Market segmentation, self-selection, and product line design. *Marketing Sci.* **3** 288–307.

Mussa, M., S. Rosen. 1978. Monopoly and product quality. *J. Econ. Theory* **18** 301–317.

Nanda, D. 1995. Strategic impact of just-in-time manufacturing on product market competitiveness. Working paper, University of Rochester, Rochester, NY.

Pine, B. 1993. *Mass Customization: The New Frontier in Business Competition.* Harvard Business School Press, Boston, MA.

Quelch, J., D. Kenny. 1994. Extend profits, not product lines. *Harvard Bus. Rev.* 153–160.

Smith, W. 1956. Product differentiation and market segmentation as alternative marketing strategies. *J. Marketing* **21** 3–8.

Tayur, S., R. Ganeshan, M. Magazine. 1998. *Quantitative Models for Supply Chain Management.* Kluwer Academic Publishers, the Netherlands.

Zipkin, P. 2000. *Foundations of Inventory Management.* McGraw-Hill, New York.

The consulting Senior Editor for this manuscript was Hau Lee. This manuscript was received on February 19, 1999, and was with the author 264 days for 3 revisions. The average review cycle time was 75 days.