2006

Dynamic Pricing through Discounts for Optimizing Multiple Class Demand Fulfillment

Qing DING
Singapore Management University, dingqing@smu.edu.sg

Panos Kouvelis
Washington University in St Louis

Joseph M. Milner
University of Toronto
DOI: https://doi.org/10.1287/opre.1060.0248

OPERATIONS RESEARCH
Vol. 54, No. 1, January–February 2006, pp. 169–183
ISSN 0030-364X | EISSN 1526-5463 | 06 | 5401 | 0169

informs

DOI 10.1287/opre.1060.0248

Dynamic Pricing Through Discounts for Optimizing Multiple-Class Demand Fulfillment

Qing Ding
School of Business, Singapore Management University, 469 Bukit Timah Road, Singapore 259756, dingqing@smu.edu.sg

Panos Kouvelis
John M. Olin School of Business, Washington University, St. Louis, Missouri 63130, kouvelis@wustl.edu

Joseph M. Milner
Joseph L. Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario, Canada M5S 3E6, milner@rotman.utoronto.ca

In a multiple-customer-class model of demand fulfillment for a single item, we consider the use of dynamic price discounts to encourage backlogging of demand for customer classes denied immediate service. Customers are assumed to arrive over several stages in a period, and customer classes are distinguished by their contractual price and sensitivity to discounts. Through dynamic programming we determine the optimal discounts to offer, assuming a linear model for the sensitivity of customers to such inducements. We show that customers are served in class order, and allocation of inventory to demand is determined by considering the current number of customers backlogged, as well as the current inventory position. Through comparison to a naive supplier allocating inventory first come/first served with no discounting, we show that profits are primarily influenced by the allocation of capacity, and the use of price discounts primarily benefits the second-class customers' overall fill rate. Heuristics for implementation of the solution in real-time settings are given.

Subject classifications: inventory/production: dynamic pricing, multiple class; dynamic programming: applications.
Area of review: Manufacturing, Service, and Supply Chain Operations.
History: Received August 2003; revisions received July 2004, December 2004; accepted December 2004.

1. Introduction

In this paper, we study the problem of allocating inventory to demand from several classes of customers when partial backlogging of unfilled demand is possible. The customer classes are distinguished by the price they pay for the item and their willingness to wait for fulfillment of demand in a subsequent period. Demand from each customer class is modeled as a realization of a (nonstationary) random variable during each of several stages into which a period is divided. The firm is able to view this demand in each stage prior to making an allocation decision on which demand to fill. Unfilled demand may then wait for later fulfillment. The probability of this occurring is influenced by a discount the firm may offer as well as some class-specific parameters. As the firm must determine how much inventory to allocate in each stage, the priority to use in serving the demand, and the discounts to offer, we refer to the problem as the allocation, discounting, and prioritization (ADP) problem.

The problem arises in a number of industries. The motivating example is based on the fulfillment of demand at a wholesaler of industrial products. At the firm's distribution center, orders are received throughout the day from customers for whom there is a fixed price for a unit (generally from contractual terms or competitive environment) and who expect same-day shipping. Given a limited inventory, the distributor may choose to offer the customer next-day shipping on the item in hope of being able to fulfill the request of a more valued customer; additional inventory is obtained overnight for the next day's demand. To induce the customer to wait for supply, the distributor may offer a discount. Similar problems are found in online and catalog businesses where firms need to determine their availability to ship on a given day. For example, an online bookseller may quote a time until shipping that is based on the ability of the firm to withdraw a unit of demand from a warehouse. Customers arriving early in the day may be quoted a longer time until shipping so that inventory may be reserved for customers coming later in the day with higher-valued orders.

This paper contributes to the literature by incorporating dynamic price discounting (i.e., offering economic incentives for customer retention fully reflecting all available inventory and realized demand information) with inventory rationing in a model that can be solved in an efficient manner. The model is complicated by the dependence of the ability to serve future demand on the discounts given to each customer class. We show that inventory should be allocated in each stage in class order as long as the inventory is above a determined threshold level. Comparing the

model to a naive first-come/first-served fulfillment without backlogging indicates that the inventory allocation and dynamic pricing serve different purposes. We observe that the increased profit of the firm vis-à-vis the naive server is primarily due to appropriate inventory allocation. Provision of discounts to the customers denied immediate service primarily ensures that overall service levels to these customers are maintained. A second benefit revealed is that the discounts provide buffer demand for subsequent stages if demand does not materialize. That is, dynamic pricing in this context is not primarily used to increase profits, but rather to provide service. We show that our solution approach can be executed with a near-optimal heuristic that can address the real-time revenue management needs of agile distribution and direct-to-market channels.

1.1. Literature Review

The research is related to work in inventory rationing, customer retention, revenue management, and dynamic pricing. Early work by Topkis (1968) considered the rationing of inventory to demand from n customer classes when a period is divided into several intervals. He shows that a base-stock ordering policy is optimal and demand is fulfilled in class order as long as inventory is above a class-dependent allocation level. A model similar to Topkis' under a different operating environment has been considered by Frank et al. (2003). Cohen et al. (1988) consider an (s, S) inventory system where two classes of customers arrive, with the higher-priority customer being served first. The focus of the paper is on the determination of the reorder level s and the order-up-to value S through the development of heuristics and approximations. Deshpande et al. (2003) analyze a static threshold-based rationing policy for a continuous-review two-demand-class system with backorders. Ha (1997) considers the problem of allocating inventory to n customer classes in a make-to-stock environment where stock replenishment is explicitly modeled as a production system through an $M/E_k/1$ queue. The optimal policy is characterized by an inventory level below which production is initiated and an inventory level for each customer class above which demand will be fulfilled for the class. De Vericourt et al. (2001) investigate various stock allocation policies in make-to-stock production systems and assess the benefits of inventory-rationing policies. Gerchak et al. (1985) consider a two-class problem where the decision is when to reject lower-class customers based on the time to go, and inventory. Weatherford et al. (1993) study the determination of dynamic allocation limits in a two-class problem with the possibility of lower-class customers purchasing at a higher-class price. In all these papers, the unsatisfied demand is either entirely backlogged or entirely lost. On the other hand, we consider the partial backlogging case where price discounts are used to induce a desirable level of backlogging. In a recent paper, Cattani and Souza (2002) investigate inventory-rationing policies in multiclass a priori determined fixed-price environments with application for firms operating in a direct market channel. They compare the performance of these rationing policies with a pure first-come, first-serve policy under various scenarios for customer response to delay (lost sales, backlog, and a combination of lost sales and backlog). The inventory system is supplied by a production system or colocated supplier with exponentially distributed processing times. Our paper emphasizes the use of dynamic policies in combination with inventory rationing in partial backlogging environments, with the backlog level affected by the dynamic pricing policies of the firm.

Another related stream of work in the inventory management literature looks at economic incentives to retain customers in the presence of stockouts. Cheung (1998) considers a continuous-review model where a discount can be offered to the customers willing to accept backorders even before the inventory is depleted, but the proportion backlogged is not a function of monetary incentives, as is the case in our work. The optimality of offering incentives to backlog demand for a simple inventory system is explored in DeCroix and Arreola-Risa (1998), but their analysis does not exploit different customer classes or dynamic discount adjustments. Further, their analysis seems to imply that the majority of cost savings result from incentives offered after a stockout occurs, which is not necessarily true in a multiclass environment such as the one examined in our paper.

Chen (2001), inspired by e-retailing environments, studies optimal pricing-replenishment strategies that balance the costs due to discounted prices and the benefits due to advance demand information from customers willing to accept longer lead times for the right discount. In the paper, the firm offers a menu of price and lead-time combinations, and customers can reveal their priorities. Chen focuses on finding an optimal menu of static prices, as opposed to dynamic discounts we use, and he does not consider inventory-rationing policies, which are essential elements of our formulation. Wang et al. (2002) study a problem of meeting demand from two demand classes with different lead-time requirements. However, they study the required inventory levels in a two-echelon supply chain, where each location follows a base-stock policy with no inventory rationing, whereas we specifically focus on dynamic pricing and inventory rationing for a single location.

The problem is also related to the well-studied revenue management problem, in which a firm seeks to determine the number of units of capacity to reserve for sale to customers arriving at a later time. Belobaba (1987) considered a heuristic approach to solving the multiple fare-class problem. Wollmer (1992), Brumelle and McGill (1993), and Robinson (1995) considered extensions and refinements determining optimal solutions. In all these models, fare classes are assumed to arrive sequentially so that the solution of the problem requires determining how much to reserve for other fare classes. A review of the extensive revenue management literature is provided by McGill

and van Ryzin (1999). In our paper, we assume concurrent arrivals of demand from different classes. A number of recent papers study a dynamic version of the perishable asset revenue management problem, where the selling price may be varied continuously over time. Gallego and van Ryzin (1994) consider a model where the demand rate for an item depends on the current price offered, and solve for an expected revenue-maximizing policy. Bitran and Mondschein (1997) study a similar problem with demand varying over time and demonstrate the effectiveness of restricting the number of prices to a small set and enforcing policies that are monotonic with respect to the price. Zhao and Zheng (2000) study an extension of Gallego and van Ryzin (1994) with nonhomogeneous reservation prices. Finally, Feng and Xiao (2000) find an exact solution for the multiple-price model in continuous time when monotonic pricing policies are assumed. Our work is related to this work as we search for allocation policies in a multiple-price model, where we explicitly incorporate the time dimension. However, our model differs from this work in that we offer dynamically adjusted price discounts only to the customers denied immediate service. Further, while previous papers have assumed rejected customers are lost, we allow for rejected customers to wait, i.e., they are partially backlogged with a probability dependent on a price discount.

Revenue management has also been studied within the economics literature. Gale and Holmes (1993) discuss the use of advance-purchase discounted prices to divert demand from a peak period to an off-peak period. They show that by doing so, a monopoly airline would expand output and total surplus. Dana (1999) presents a model in which the airline does not know when the peak demand will occur and shows that by offering multiple prices and rationing inventory, the airline is able to shift demand from a peak to an off-peak period. Similar to our model, customers incur a waiting cost for being served in their nonpreferred period. Both models consider two classes of customer and stationary pricing, in contrast to our model of dynamically adjusted discounts.

Outside of the revenue management context, there has been work on integrating dynamic pricing with production/supply policy, but mostly for single-product homogeneous customer populations (e.g., Zabel 1972, Thowsen 1975). In more recent work, Chan et al. (2005) consider a multiperiod deterministic demand model where pricing and production decisions must be made for each period with some capacity constraint. Similarly, Federgruen and Heching (1999), in a stochastic demand setting, focus on showing the optimality of a base-stock/list-price policy.

1.2. Paper Organization

The remainder of this paper is organized as follows. In §2, we formally introduce the model. We present the solution to the problem in §3. In §4, we discuss properties of the ADP optimal policy and provide useful insights from some numerical experiments. In §5, we consider two heuristics. One discusses the use of the expected demand in solving the problem; the second discusses a continuous-time version of the problem. We draw our conclusions in §6.

2. Model

Consider a model where, at the start of a given period, inventory is purchased up to some base-stock level. The period is subsequently divided into M stages. For example, a period may represent a day and each stage may represent an hour. In each stage $j \in \{1, \ldots, M\}$, demand from each of the K customer classes is realized. The firm then determines which customer demands to fill. If a customer's order is not filled immediately, the firm offers a discount to encourage the customer to agree to wait for fulfillment in the next period. The firm may also decide to fill demand from waiting customers (those denied service in a previous stage), rather than delaying their service until the subsequent period. Each customer class $i \in I = \{1, \ldots, K\}$ is distinguished by a contractual price per unit, p_i, and parameters α_{ij} and β_{ij} that characterize the customers' willingness to wait for delayed delivery of their demand. We assume, without loss of generality, that $p_1 \leqslant \cdots \leqslant p_K$. Let d_{ij} be the demand from customer class i in stage j. We assume that d_{ij} is a discrete, nonnegative, independent (by customer type and stage) random variable, the distribution of which is known. For simplicity, we assume that each customer orders exactly one unit. After demand in a stage is revealed, the firm determines the amount of inventory to allocate, which customers to assign the allocated inventory, and what discounts to offer to customers not allocated inventory to induce them to wait for supply in the next period. Inventory held at the end of the last stage is carried forward to the next period, and waiting demand is served at the start of the next period. Let X_j be the initial inventory of the firm in stage j and let Y_j be the amount of inventory allocated in stage j to demand. (Table 1 provides a glossary of the notation used.)

We adopt the following costs. The per-unit cost to the firm for the inventory is c_p. If the number of units allocated in stage j is larger than a given capacity, g_j, the firm incurs an additional cost per unit, c_j, which we refer to as a congestion cost. This cost results from an unplanned overload of resources and is included to model the operational cost to the firm of delaying fulfillment of demand until a later stage. We allow partial backlogging of demand. The firm incurs a cost c_w per unit (the waiting cost) for demand that is backlogged and delivered in the next period. The firm incurs a cost c_l per unit for demand that is lost. We let h be the holding cost per unit held at the end of the period.

We distinguish between new customers in stage j, i.e., customers whose demand occurs in stage j, and waiting customers, i.e., customers whose demand was not fulfilled in a prior stage and are waiting for demand fulfillment, either later in the current period or in the next period. Let z_{ij} be the discount offered to new customers of class i

Table 1. Notation.

p_i	price per unit to customer class i
c_p	cost per unit to the firm
c_j	congestion cost per unit in stage j
g_j	firm capacity in stage j
c_w	waiting cost per unit
c_l	lost-sales cost per unit (in addition to lost revenue)
h	holding cost per unit per unit time
z_{ij}	discount per unit to class i in stage j
α_{ij}	probability that a class i customer will wait in stage j independent of discount
β_{ij}	linear coefficient for a discount z_{ij}
γ_{ij}	probability of waiting
X_j	inventory at start of stage j
Y_j	inventory allocated in stage j
d_{ij}	new demand from class i in stage j
$\vec{d}_{ij}$	waiting demand from class i in stage j
$\bar{d}_j$	vector of demand in stage j
D_{ij}	new demand from class i and higher in stage j
$\vec{D}_j$	number of waiting customers at start of stage j
$\bar{D}_j$	total demand in stage j
ω_j	sequence of demand fulfillment in stage j
$D^{\omega}_{lij}(\vec{D}^{\omega}_{lij})$	position in sequence ω_j that the lth new (waiting) customer of class i is served in stage j
$\pi_{lij}(\vec{\pi}_{lij})$	contribution to the profit for serving the lth new (waiting) customer of class i in stage j
π_j	expected profit from stage j onward given z_j, ω_j, and Y_j
Π_j	optimal expected profit from stage j onward
$\Delta_X\Pi_j$	change in profit if an additional unit of inventory is available in stage j
$\Delta_{\vec{D}}\Pi_j$	change in profit if an additional customer is waiting in stage j
L_{ij}	value of serving an additional customer of class i in stage j

in stage j to induce them to wait for delayed fulfillment and let $z_j = \{z_{ij}\}_{i\in I}$. If the customer agrees to wait, the price for the unit of demand is $p_i - z_{ij}$. We assume that the customer commits to paying this discounted price and that the firm commits to delivering the unit either in a subsequent stage in the current period or at the start of the next period at this price. No additional incentive is given to waiting customers in a stage j.

Let the probability that a customer will wait be γ_{ij}, and we assume for simplicity that it is defined by a linear function

$$\gamma_{ij} = \alpha_{ij} + \beta_{ij} z_{ij},$$

that is, those customers not served choose to wait based on independent Bernoulli random variables. We assume that α_{ij} and β_{ij} are nonincreasing in i, i.e., $0 \leqslant \alpha_{Kj} \leqslant \cdots \leqslant \alpha_{1j} \leqslant 1$ and $0 \leqslant \beta_{Kj} \leqslant \cdots \leqslant \beta_{1j}$. This restriction implies that customer classes with higher contractual prices are less willing to wait for delayed service (based on the lower α_{ij} value) and are also less likely to respond to a discount (based on the lower β_{ij} value). Further, we restrict $0 \leqslant z_{ij} \leqslant (1-\alpha_{ij})/\beta_{ij}$ so that $\alpha_{ij} \leqslant \gamma_{ij} \leqslant 1$. We justify these assumptions by considering the firm to be engaged with their customers through contractual agreements, promising higher service levels for customers willing to pay higher contracted prices. We assume that the firm must offer greater discounts to such customers to retain their business. Similarly, customers paying lower contracted prices do so knowing that their service is more likely to be delayed.

We note that these assumptions are limiting in that they imply a negative correlation between the customer's willingness to pay and his/her willingness to wait that may not be natural in all environments. For example, Alderman (1987) shows that rationing through queueing in a developing economy creates a two-part tariff that benefits those more able to pay higher costs. That is, a customer whose valuation of a good is higher may be more willing to wait. Theoretical justification for this is discussed in Barzel (1974). However, in other contexts, such as at gasoline service stations, the willingness to wait has been observed to be negatively correlated with willingness to pay (e.g., Deacon and Sonstelie 1985, Png and Reitman 1994). Thus, alternate models of the relationship between valuation and willingness to wait are justified, and relaxing these assumptions should provide interesting insights. A similar relaxation is considered by Bruce et al. (2004), which studies how independence between willingness to pay and ability to pay for durable goods affects promotions to consumers. However, such a relaxation is beyond the scope of this paper.

We make several assumptions regarding the costs to avoid trivial or undefined solutions. We assume that $(1-\alpha_{ij})/\beta_{ij} \leqslant p_i + c_l$ so that the discount does not exceed the economic value of a unit of inventory. We assume that $p_i + c_l - c_p - c_w \geqslant 0$ and $c_p \geqslant h$ so that the firm prefers backlogged customers to lost customers and holds excess inventory. Finally, we assume that $c_w + h \geqslant c_M$ so that serving a waiting customer at the end of a period is preferable to delaying the customer until the next period while holding inventory.

Let $\vec{d}_{ij}$ be the r.v. for the number of customers from class i waiting from a previous stage in stage j.

Let $\bar{d}_j$ be the random vector of all demand (new and waiting) in stage j;

$\bar{d}_j = \{d_{1j}, \ldots, d_{Kj}, \vec{d}_{1j}, \ldots, \vec{d}_{Kj}\}$.

(We use the notational convention that symbols with an arrow, e.g., $\vec{d}_{ij}$, refer to demand from class i waiting from a previous stage, symbols with a bar, e.g., $\bar{d}_{ij}$, refer to both new and waiting demand, and unaccented symbols, e.g., d_{ij}, refer to new demand only.)

Let $D_{ij} = \sum_{k=i}^{k=K} d_{kj}$ be the new demand from classes i through K in stage j, let $\vec{D}_j = \sum_{i\in I} \vec{d}_{ij}$ be the number of waiting customers at the start of stage j, and let $\bar{D}_j = \sum_{k=1}^{k=K} d_{kj} + \vec{D}_j$ be the total demand in stage j.

We next consider the order in which customers are served in a period. Let Ω_j be the set of all permutations

of $(1, \ldots, \bar{D}_j)$ and let $\omega_j \in \Omega_j$. Then, each ω_j corresponds to an ordering of the customers as follows. Let D^{ω}_{lij} be the position in the service order ω_j for the lth new customer from class i in stage j and let $\vec{D}^{\omega}_{lij}$ be the position in the order for the lth waiting customer from class i. (Example: Suppose that there are two classes with two new customer demands each in stage 1. Then, $\bar{D}_1 = 4$. Suppose that $\omega_j = \{1, 4, 2, 3\}$, which we interpret to imply $D^{\omega}_{111} = 1$, $D^{\omega}_{211} = 4$, $D^{\omega}_{121} = 2$, and $D^{\omega}_{221} = 3$. Then, the first customer from class 1 is served first, followed by the first customer from class 2, the second customer from class 2, and finally the second customer from class 1.) As long as the subscripts of D_{lij} are consistently ordered, each ω_j corresponds to a different ordering or prioritization of service.

At the start of each stage j in a period, the state is described by the pair $(X_j, \bar{d}_j)$. Let $\pi_{lij}(z_{ij}, \omega_j, Y_j \mid X_j, \bar{d}_j)$ be the profit contribution received from the lth unit of demand from new class i customers in stage j given the discounts, prioritization, and allocation. Considering the cost of previously purchased inventory a sunk cost, $\pi_{lij} = p_i$ if the demand is served. If the demand is not served and the customer waits, $\pi_{lij} = p_i - z_{ij} - c_p - c_w$, reflecting the discounted revenue less the cost of purchasing a unit of inventory to fulfill the waiting demand in the subsequent period less the cost of delaying the customer. Finally, if the demand is lost, $\pi_{lij} = -c_l$. Thus, the expected contribution is

$$\pi_{lij}(z_{ij}, \omega_j, Y_j \mid X_j, \bar{d}_j) = \begin{cases} p_i, & Y_j \geqslant D^{\omega}_{lij}, \\ \gamma_{ij}(p_i - z_{ij} - c_p - c_w) - (1 - \gamma_{ij})c_l, & Y_j < D^{\omega}_{lij}. \end{cases} \quad (1)$$

Similarly, let $\vec{\pi}_{lij}(z_{ij}, \omega_j, Y_j \mid X_j, \bar{d}_j)$ be the contribution to the profit for serving the lth waiting customer from class i in stage j. From (1), observe that the firm incurs the unit production cost and the cost of delaying the customer until the next period when the customer is first delayed. If the customer is served in stage j, these costs must be charged back. Thus,

$$\vec{\pi}_{lij}(z_{ij}, \omega_j, Y_j \mid X_j, \bar{d}_j) = \begin{cases} c_p + c_w, & Y_j \geqslant \vec{D}^{\omega}_{lij}, \\ 0, & Y_j < \vec{D}^{\omega}_{lij}. \end{cases} \quad (2)$$

Let $\Pi_j(X_j, \bar{d}_j)$ be the optimal expected profit from stage j onward given the current state. At the end of stage M, any remaining inventory is carried forward to the next period. Recalling that the policy of the firm is to purchase up to a given base-stock level at the start of each period, each unit of inventory is valued at the production cost less the holding cost. Therefore, we define

$$\Pi_{M+1}(X_{M+1}, \bar{d}_{M+1}) = (c_p - h)X_{M+1}. \quad (3)$$

Similarly, we assume that any customer's demand that is not fulfilled within the period is served first at the start of the next period with units purchased at the cost c_p. Note that we assume that the contracted prices are firm and no renegotiation is considered at the start of the next period, i.e., delayed customers always pay $p_i - z_{ij}$ and the firm always serves them prior to serving any new demand in the next period.

Let $\pi_j(z_j, \omega_j, Y_j \mid X_j, \bar{d}_j)$ be the expected profit per stage from stage j onward for a given z_j, Y_j, ω_j and state $(X_j, \bar{d}_j)$. Then,

$$\pi_j(z_j, Y_j, \omega_j \mid X_j, \bar{d}_j) = \sum_{i=1}^{K}\left(\sum_{l=1}^{d_{ij}} \pi_{lij}(z_j, \omega_j, Y_j \mid X_j, \bar{d}_j) + \sum_{l=1}^{\vec{d}_{ij}} \vec{\pi}_{lij}(z_j, \omega_j, Y_j \mid X_j, \bar{d}_j)\right) - c_j(Y_j - g_j)^+ + E_{\bar{d}_{j+1}}[\Pi_{j+1}(X_{j+1}, \bar{d}_{j+1})], \quad (4)$$

where X_{j+1} is the inventory at the start of stage $j+1$ and the state $\bar{d}_{j+1}$ reflects the new customers in stage $j+1$ and the waiting customers not served in stage j. That is, $d_{i,j+1}$ is the random number of new customers and $\vec{d}_{i,j+1}$ is a shifted binomial random variable given as follows: Let $A^{\omega}_{ij} = \vec{d}_{ij} - \sum_{l=1}^{\vec{d}_{ij}} 1_{\vec{D}^{\omega}_{lij} \leqslant Y_j}$ represent the number of waiting customers in period j still waiting in period $j+1$ and let $B^{\omega}_{ij} = A^{\omega}_{ij} + d_{ij} - \sum_{l=1}^{d_{ij}} 1_{D^{\omega}_{lij} \leqslant Y_j}$ be the total possible number of waiting customers in stage $j+1$. Then,

$$P(\vec{d}_{i,j+1} = d) = \text{binomial}(d - A; B - A, \gamma_{ij}) \quad (5)$$

for $d = A \cdots A + B$, where binomial $(x; n, p) = \binom{n}{x} p^x \cdot (1-p)^{n-x}$.

(Note that $x^+ = \max[x, 0]$ and we let the indicator function $1_x = 1$ if x is true, and 0 otherwise.)

Noting our assumptions that $c_p > h$, it is clear in any optimal solution that

$$X_{j+1} = X_j - Y_j. \quad (6)$$

Equations (5) and (6) define the system state constraints for the following dynamic program.

The Allocation, Discounting, and Prioritization (ADP) Problem

The optimal profit for stage j onwards is given as

$$\Pi_j(X_j, \bar{d}_j) = \max_{z_j, \omega_j, Y_j} \pi_j(z_j, \omega_j, Y_j \mid X_j, \bar{d}_j) \quad (7a)$$

subject to

$$0 \leqslant z_{ij} \leqslant \frac{1 - \alpha_{ij}}{\beta_{ij}} \quad \text{for all } i \in I, \quad (7b)$$

$$\omega_j \in \Omega_j, \quad (7c)$$

$$0 \leqslant Y_j \leqslant \min[X_j, \bar{D}_j] \quad \text{and} \quad Y_j \text{ integer}. \quad (7d)$$

The problem at the start of each period is then to determine the initial inventory level, X_1, prior to observing demand in the first stage,

$$\max_{X \geqslant 0} -c_p X + E_{\bar{d}_1}[\Pi_1(X, \bar{d}_1)].$$

3. ADP Problem Solution

We solve the problem through dynamic programming, starting first with the final stage in a period, M, and then solving the problem by induction for stages $M-1,\ldots,1$. In each stage, three decisions need to be made: what discounts to offer, what prioritization to use, and finally, how much inventory to allocate to demand in the stage. Let $z_j^* = \{z_{ij}^*\}_{i\in I}$ be the set of optimal discounts, let ω_j^* be the optimal priority sequence, and let Y_j^* be the optimal allocation in stage j.

3.1. Stage M

The following proposition summarizes the result for the final stage.

PROPOSITION 1. *Let*

$$z'_{iM} = \frac{p_i + c_l - c_p - c_w}{2} - \frac{\alpha_{iM}}{2\beta_{iM}} \quad \textit{for each } i \in I.$$

Then,

(a) *The unique optimal price discount is*

$$z_{iM}^* = \left(\min\left[z'_{iM}, \frac{1-\alpha_{iM}}{\beta_{iM}}\right]\right)^+ \quad \textit{for each } i \in I;$$

(b) *A priority sequence that serves new customers, d_M, in class order from K to 1 prior to serving any waiting demand, $\vec{d}_M$, is optimal; and*

(c) *All customers, new and waiting, should be allocated inventory if possible, i.e., $Y_M^* = \min[X_M, \bar{D}_M]$.*

Comments. Observe that the discount offered can be interpreted as the profit maximizing discount of a monopolist where the demand is increasing linearly in the discount $(\alpha_{iM} + \beta_{iM} z_{iM})$ and the revenue $(p_i - z_{iM})$ is decreasing in the discount, subject to a nonnegativity constraint and a constraint on the total probability being less than 1. As such, the discount, z_{iM}^*, does not depend on X_M, Y_M, $\bar{d}_M$, or ω_M. This allows us to determine the prioritization, ω_M, and the allocation of Y_M, easily.

Observe also that there is no priority given to waiting customers based on their class. This follows from the profit margins given in (2), which recognizes that the change in profit for allocating inventory to a waiting customer is $c_p + c_w$, independent of class because this is the charge back to the firm for serving a delayed customer in a period. Because waiting customers receive the lowest priority, we therefore denote the waiting customers by class 0 and use this designation in subsequent discussions, where convenient, rather than the notation $\vec{\cdot}$.

Also note from Proposition 1(a) that $z_{iM}^* < p_i + c_l$. Such a result makes economic sense, as $p_i + c_l$ represents the total cost of lost demand.

To show the analogous results for the general stage j, we need to prove several monotonicity properties. To do so we introduce the following notation: Let $\Delta_X \Pi_j(X_j, \bar{d}_j)$ be the change in the profit in stage j if an additional unit of inventory is available for allocation in stage j, i.e., if $X_j + 1$ units are available rather than X_j units. Let $\Delta_{\bar{D}} \Pi_j(X_j, \bar{d}_j)$ be the change in profit if an additional customer is waiting in stage j, i.e., if the number waiting is $\vec{D}_j + 1$ rather than $\vec{D}_j$.

Note from (3) that $\Delta_X \Pi_{M+1}(X_{M+1}, \bar{d}_{M+1}) = c_p - h$, and also note from the discussion above that $\Delta_{\bar{D}} \Pi_{M+1}(X_{M+1}, \bar{d}_{M+1}) = 0$.

Given the set of optimal discounts $\{z_{ij}^*\}_{i\in I}$, let $\gamma_{ij}^* = \alpha_{ij} + \beta_{ij} z_{ij}^*$ for each i and j be the corresponding probability that a customer waits. Let

$$L_{ij}(z_{ij}^*, \omega_j^*, Y_j^* \mid X_j, \bar{d}_j) = p_i + c_l - c_j \cdot 1_{Y_j^* \geq g_j} - \gamma_{ij}^*\left(p_i + c_l - z_{ij}^* - c_p - c_w + E_{\bar{d}_{j+1}}[\Delta_{\bar{D}} \Pi_{j+1}(X_j - Y_j^*, \bar{d}_{j+1})]\right)$$

for $i = 1,\ldots,K$, and let

$$L_{0,j}(z_{0,j}^*, \omega_j^*, Y_j^* \mid X_j, \bar{d}_j) = c_p + c_w - c_j \cdot 1_{Y_j^* \geq g_j} - E_{\bar{d}_{j+1}}[\Delta_{\bar{D}}(X_j - Y_j, \bar{d}_{j+1})].$$

From (1) and (2), we observe that L_{ij} represents the value of serving an additional customer from class i, $i = 0,\ldots,K$, in stage j.

PROPOSITION 2. (a) $E_{\vec{d}_M}[\Delta_X \Pi_M] \geq c_p - h$. $E_{\vec{d}_M}[\Delta_X \Pi_M]$ *is nonincreasing in X_M and nondecreasing in $\vec{D}_M$.*

(b) $0 \leq E_{\vec{d}_M}[\Delta_{\bar{D}} \Pi_M] \leq c_p + c_w$. $E_{\vec{d}_M}[\Delta_{\bar{D}} \Pi_M]$ *is nondecreasing in X_M and nonincreasing in $\vec{D}_M$.*

Comments. The proposition implies that the profit increases with diminishing returns as the inventory and number of waiting customers in period M increases. Also, the value of additional inventory increases as the number of waiting customers increases, and the value of waiting customers increases as the amount of inventory increases.

3.2. Stage j

We now consider the problem faced in stage j, showing by induction how to determine the discounts, z_{ij}, the prioritization, ω_j, and the inventory to allocate, Y_j. We present an algorithm to determine the discounts. We then show (as in the case of stage M) that the optimal priority sequence assigns inventory to new customers in class order prior to serving any waiting demand. Finally, we show how to determine the amount of inventory to allocate in stage j.

We assume for the induction that for stages $k > j$, $E_{\vec{d}}[\Delta_X \Pi_k] \geq c_p - h$, nonincreasing in X_k and nondecreasing in $\vec{D}_k$; and $0 \leq E_{\vec{d}}[\Delta_{\bar{D}} \Pi_k] \leq c_p + c_w$, $E_{\vec{d}}[\Delta_{\bar{D}} \Pi_k]$, nonincreasing in $\vec{D}_k$ and nondecreasing in X_k.

Determining the Discount z_{ij}^*. Let

$$z'_{ij} = \frac{p_i + c_l - c_p - c_w}{2} - \frac{\alpha_{ij}}{2\beta_{ij}} + \frac{1}{2}E_{\vec{d}}[\Delta_{\vec{D}}\Pi_{j+1}(X_j - Y_j, \vec{d}_{j+1})]. \quad (8)$$

The optimal discount z_{ij} given Y_j and ω_j solves

$$\max \ \Pi_j(z_j, \omega_j, Y_j \mid X_j, \vec{d}_j) \quad (9a)$$

$$\text{subject to } 0 \leqslant z_{ij} \leqslant \frac{1-\alpha_{ij}}{\beta_{ij}} \quad \text{for all } i \in I. \quad (9b)$$

Consider the unconstrained problem

$$\max \Pi_j(z_j, \omega_j, Y_j \mid X_j, \vec{d}_j).$$

The first-order conditions for each $i \in I$ are

$$\begin{aligned}\frac{\partial \Pi_j}{\partial z_{ij}} &= \sum_{l=1}^{d_{ij}} \frac{\partial \pi_{lij}}{\partial z_{ij}} + \frac{\partial}{\partial z_{ij}} E[\Pi_{j+1}(z_{ij}, Y_j, \omega_j)] \\ &= \frac{\partial}{\partial z_{ij}} \sum_{l \text{ s.t. } Y_j < D_{lij}} \pi_{lij} + \frac{\partial}{\partial z_{ij}} \gamma_{ij} E[\Delta_{\vec{D}}\Pi_{j+1}] \\ &= \frac{\partial}{\partial z_{ij}} \sum_{l \text{ s.t. } Y_j < D_{lij}} \gamma_{ij}(p_i + c_l - z_{ij} - c_p \\ &\qquad - c_w + E[\Delta_{\vec{D}}\Pi_{j+1}]) \\ &= \sum_{l \text{ s.t. } Y_j < D_{lij}} -\alpha_{ij} - 2\beta_{ij} z_{ij} \\ &\quad + \beta_{ij}(p_i + c_l - c_p - c_w + E[\Delta_{\vec{D}}\Pi_{j+1}]) \\ &= 0\end{aligned}$$

for each $i \in I$. Based on the first-order conditions, the solution to the unconstrained problem is given as z'_{ij} above. Note that the second equality holds, as changing z_{ij} changes the probability that the lth customer waits, the economic value of which is $E[\Delta_{\vec{D}}\Pi_{j+1}]$. Because $E[\Delta_{\vec{D}}\Pi_{j+1}]$ is nonincreasing in $\vec{D}$, and the probability that a customer waits, γ_{ij}, is increasing in z_{ij}, $\partial^2\Pi_j/\partial z_{ij}^2 < 0$. Thus, Π_j is concave in z_j and we have

PROPOSITION 3. *$z_{ij}^* = 0$, z'_{ij} or $(1-\alpha_{ij})/\beta_{ij}$ for all i and j.*

The difficulty is that z'_{ij} depends on $E[\Delta_{\vec{D}}\Pi_{j+1}(X_j - Y_j, \vec{d}_{j+1})]$ where the expectation is taken over d_{j+1} and $\vec{d}_{j+1}$, the latter of which is dependent on γ_{ij} through (5) and therefore dependent on the value of z_{ij}. Thus, to calculate $E[\Delta_{\vec{D}}\Pi_{j+1}]$, we need to know z_{ij}. However, we know by the Kuhn-Tucker conditions for (9a) and (9b), that for classes x and y if $0 < z_{xj}^* < (1-\alpha_{xj})/\beta_{xj}$ and $0 < z_{yj}^* < (1-\alpha_{yj})/\beta_{yj}$, then

$$\frac{\partial \Pi_j}{\partial z_{xj}} = \frac{\partial \Pi_j}{\partial z_{yj}} = 0,$$

so that

$$(\alpha_{xj} - 2\beta_{xj}z_{xj}^* + \beta_{xj}p_j) - (\alpha_{yj} - 2\beta_{yj}z_{yj}^* + \beta_{yj}p_j) = 0$$

or

$$\Delta z_{xy} \equiv z_{xj}^* - z_{yj}^* = \frac{1}{2}\left(\left(p_x - \frac{\alpha_{xj}}{\beta_{xj}}\right) - \left(p_y - \frac{\alpha_{yj}}{\beta_{yj}}\right)\right).$$

This implies that all classes x, y not at their bounds must be separated by Δz_{xy}. Further, this gives us a procedure to calculate z_{ij} for all i. We do so by first ordering the classes by the value of $p_i - \alpha_{ij}/\beta_{ij}$ in decreasing order and setting $z_{ij} = 0$ for all i. We then iteratively increase the values of z_{ij} for classes in the given order, maintaining a separation of z_{xy} for all classes x, y not at their upper bounds. Because in any iteration we know the values of z_{ij} for all i, we can calculate $E[\Pi_{j+1}(X - Y, d)]$ for a given ω_j and Y_j and thus find $E[\Delta_{\vec{D}}\Pi_{j+1}]$. We can then test if (8) holds for those classes not at their boundary. The following algorithm formalizes this procedure.

Algorithm 1: Determining the Discount z_{ij}

Initialization. Let $r_i = p_i - \alpha_i/\beta_i$ and $b_i = (1-\alpha_i)/\beta_i$. (Throughout the algorithm we suppress the stage subscript j for ease of presentation.) Order the classes by r_i. (For purposes of clarity, we assume, without loss of generality, that $r_1 \geqslant \cdots \geqslant r_K$.) Set $z_i = 0$ for all $i \in I$. Set $x = 1$ (x is the last class entering the active set, i.e., the set of classes for which z_i is not at its upper or lower bounds). Let $A = \{1\}$ be the set of customer classes that are active (including a class entering the active set). Set $\Delta z_x \equiv (r_x - r_{x+1})/2$ for $1 \leqslant x < K$ and $\Delta z_K = \infty$.

Step 1. Repeat $A := A \cup \{x+1\}$, while $\Delta z_x = 0$. Set $\Delta b = \min_{y \in A}(b_y - z_y)$. Let $B = \{y \mid b_y - z_y = \Delta b\}$ be the set of active classes that would next achieve their upper bounds. Set $\Delta z = \min[\Delta b, \Delta z_x]$.

Step 2. Test: Is there $\Delta z' \in (0, \Delta z)$ such that by letting $z'_y = z_y + \Delta z'$ for all $y \in A$, z'_y solves (8) for one class in A. If so, let $z_y := z_y + \Delta z'$ for all $i \in A$ and stop. (The solution is optimal at the current levels of z_y.) Otherwise, go to Step 3.

Step 3. $\Delta z_x := \Delta z_x - \Delta b$. $z_y := z_y + \Delta z$ for all $y \in A$. Set $A := A \backslash B$ if $\Delta z = \Delta b$.

Step 4. If $A = \varnothing$ and $x = K$, stop. (All z_x are at their upper bounds.)

Step 5. If $A \neq \varnothing$ and $\Delta z_x > 0$, go to Step 1.

Step 6. $A := A \cup \{x+1\}$. $x := x+1$. Go to Step 1.

The algorithm requires at most $2K$ iterations as either one class enters or at least one class exits the active set A in any iteration. Each iteration requires $O(\log(p_1 - \alpha_1/\beta_1))$ tests of (8) because the change in z in an iteration, Δz, is bounded by $(p_1 - \alpha_1/\beta_1)$; and the concavity of $E[\Delta_{\vec{D}}\Pi]$ is given by the induction assumption.

Determining the Optimal Prioritization ω_j^*. As in the case for stage M the optimal prioritization is as follows.

PROPOSITION 4. *A priority sequence that serves new customers, d_j, in class order from K to 1 prior to serving any waiting demand, $\bar{d}_j$, is optimal.*

The proof of the proposition relies on the assumption that the classes with the highest contracted per-unit prices, p_i, are least willing to wait for delayed service and are least sensitive to any discounts offered. For those customers that have already been delayed service, giving them priority over another customer would lower the profit, as they already have received a discount for waiting, while delaying a new customer would imply such a discount would be incurred again. Also note that there is no prioritization over waiting customers as every such customer served has the same value, $c_p + c_w - E[\Delta_{\bar{D}}\Pi]$, which represents the savings in the unit cost and the waiting cost less any potential value of having the customers waiting in a later stage.

Determining the Allocation Y_j^*. Let

$$\begin{aligned}
&\Delta_Y\Pi_j^i(z_{ij}^*, \omega_j^*, Y)\\
&\quad = L_{ij} - E[\Delta_X\Pi_{j+1}(X_j - Y, \bar{d}_{j+1})]\\
&\quad = (1-\gamma_{ij}^*)(p_i + c_l) + \gamma_{ij}^*(z_{ij}^* + c_p + c_w\\
&\qquad\qquad - E[\Delta_{\bar{D}}\Pi_{j+1}(X_j - Y, \bar{d}_{j+1})]) - c_j \cdot 1_{Y \geqslant g_j}\\
&\qquad - E[\Delta_X\Pi_{j+1}(X_j - Y, \bar{d}_{j+1})]. \qquad (10)
\end{aligned}$$

$\Delta_Y\Pi_j^i$ is a well-defined function of Y. Note though, from (4), that $\Delta_Y\Pi_j^i$ is the increase in profit if an additional unit of inventory in excess of Y units is allocated to period j and if the customer that receives the unit is from class i, $i \in I$. That is,

$$\Delta_Y\Pi_j^i(z_j^*, \omega_j^*, Y) = \Pi_j(z_{ij}^*, \omega_j^*, Y_{ij} + 1) - \Pi_j(z_{ij}^*, \omega_j^*, Y_{ij})$$

if for some $l \in [1, \ldots, d_{ij}]$, $D_{lij}^{\omega^*} = Y + 1$.

Observe that

$$\begin{aligned}
&\Delta_Y\Pi_j^i(z_{ij}^*, \omega_j^*, Y) - \Delta_Y\Pi_j^i(z_{ij}^*, \omega_j^*, Y-1)\\
&\quad = \gamma_{ij}(E[\Delta_{\bar{D}}\Pi_{j+1}(X_j - Y + 1, \bar{d}_{j+1})]\\
&\qquad\qquad - E[\Delta_{\bar{D}}\Pi_{j+1}(X_j - Y, \bar{d}_{j+1})])\\
&\qquad - (E[\Delta_X\Pi_{j+1}(X_j - Y, \bar{d}_{j+1})]\\
&\qquad\qquad - E[\Delta_X\Pi_{j+1}(X_j - Y + 1, \bar{d}_{j+1})]) - c_j \cdot 1_{Y-1=g_j}\\
&\quad \leqslant E[\Delta_{\bar{D}}\Pi_{j+1}(X_j - Y + 1, \bar{d}_{j+1})\\
&\qquad + \Delta_X\Pi_{j+1}(X_j - Y + 1, \bar{d}_{j+1})]\\
&\qquad - E[\Delta_{\bar{D}}\Pi_{j+1}(X_j - Y, \bar{d}_{j+1}) + \Delta_X\Pi_{j+1}(X_j - Y, \bar{d}_{j+1})]\\
&\quad \leqslant 0.
\end{aligned}$$

The first inequality follows from $\gamma_{ij} \leqslant 1$ and $c_j \geqslant 0$. The second inequality holds because by Proposition 4 additional inventory is first used for new customers, so that each additional unit is used for a less valuable unit of demand. It follows that $\Delta_Y\Pi_j^i$ is nonincreasing in Y. Based on this, we propose the following rule.

Inventory Allocation Rule. Let Y_{ij} be the smallest value of $Y \geqslant 0$ such that $\Delta_Y\Pi_j^i < 0$. Allocate $Y_j^* = \min[Y_{ij}, X_j]$ if the Y_{ij}th customer in the optimal priority sequence ω_j^* is of class i, i.e., $D_{lij}^{\omega^*} = Y_{ij}$ for some $l \in \{1, \ldots, d_{ij}\}$. If $\bar{D}_j < Y_{0,j}$, i.e., the total demand is less than the optimal amount to allocate for all the waiting demand, then $Y_j^* = \min[\bar{D}_j, X_j]$.

The rule determines Y_{ij} by allocating inventory to stage j assuming class that i customers receive the additional inventory. It does so until allocating one unit of inventory in addition to Y_{ij} in period j lowers the firm's expected profit. By the preceding observation, Y_{ij} exists. The rule then finds the value of Y_{ij} that corresponds to a class i customer receiving the last unit. Observe that $D_{lij}^{\omega^*}$ is nonincreasing in i by Proposition 4 and Y_{ij} is nondecreasing in i because L_{ij} are nondecreasing in i. Therefore, noting the constraint $Y_j \leqslant X_j$, we have the following proposition.

PROPOSITION 5. *The optimal number of units to allocate in stage j is unique and is given by the inventory allocation rule.*

Observe that to solve for Y_j^* through the inventory allocation rule, we need to know the value of z_j^*. However, in (8), to find z_{ij}, we assumed a value of Y_j^*. Observe from (10) that Y_{ij} is nondecreasing in z_{ij}, while from (8) and the induction assumption, z_{ij} is nonincreasing in Y_{ij}. Therefore, we must iteratively solve for Y_{ij} and z_{ij} to find their optimal values. Note that constraints (7b) and (7d) are independent of each other, so the problem is separable and the iterative solution can be found readily.

Having established z_{ij}^*, ω_j^*, and Y_j^* in Propositions 3, 4, and 5, we need to show that the induction assumptions hold.

PROPOSITION 6. (a) $E_{\bar{d}}[\Delta_X\Pi_j] \geqslant c_p - h$, $E_{\bar{d}}[\Delta_X\Pi_j]$ *is nonincreasing in X_j, and nondecreasing in $\bar{D}_j$.*

(b) $0 \leqslant E[\Delta_{\bar{D}_j}\Pi_j] \leqslant c_p + c_w$, $E[\Delta_{\bar{D}_j}\Pi_j]$ *is nondecreasing in X_j, and nonincreasing in $\bar{D}_j$.*

We have therefore shown that in each stage we can determine the price discounts, the prioritization and the allocation. Recalling that at the start of each period all waiting demand is filled prior to allocating inventory to the first stage; the order-up-to level for the first stage is given by choosing the smallest X such that

$$E_{\bar{d}}[\Delta_X\Pi(X, \bar{d})] - c_p \leqslant 0.$$

4. Numerical Analysis

In this section, we study numerically the solution of the ADP problem considering comparative statics on the optimal expected profit and the optimal base-stock level. To highlight the contribution of the allocation,

prioritization, and discounting decisions, we compare the solution to the case where no discounts are given ($z = 0$) to the first-come/first-served (FCFS) case where no prioritization or allocation decisions are made, and to the FCFS case with $z = 0$—this last case representing a naive solution. Under the FCFS policy, customers are served until supply is depleted, after which discounts are given. Note that it is easy to show that the discounts in the FCFS case are given by Proposition 1(a). As in the ADP problem, we can then solve for the optimal base stock numerically (note that in some cases this may be done in closed form for the FCFS case).

EXAMPLE 1. Consider a two-stage, two-customer class problem with $p = (1.2, 2)$, $\alpha_{i1} = \alpha_{i2} = (0, 0)$, and $\beta_{i1} = \beta_{i2} = (10, 0.1)$. All customer demands d_{ij} are independent (discretized) normally distributed r.v.s with $\mu = 300$ and $\sigma = 90$. We let $c_p = 1$, $h = 0.6$, and $c_w = c_l = 0$, and assume that $c_1 = 0.6$ with $g_1 = 0$, $c_2 = 0$, and $g_2 = 0$. Note that letting $c_l = 0$ is without loss of generality, as we could equivalently let $c_l > 0$ and then use $p_i' = p_i - c_l$ throughout and obtain the same results. Similarly, letting $c_w = 0$ is also without loss of generality, as we could let $c_p' = c_p - c_w$ for some $c_w > 0$ and obtain the same results (except for the base-stock value which would change in a clear fashion).

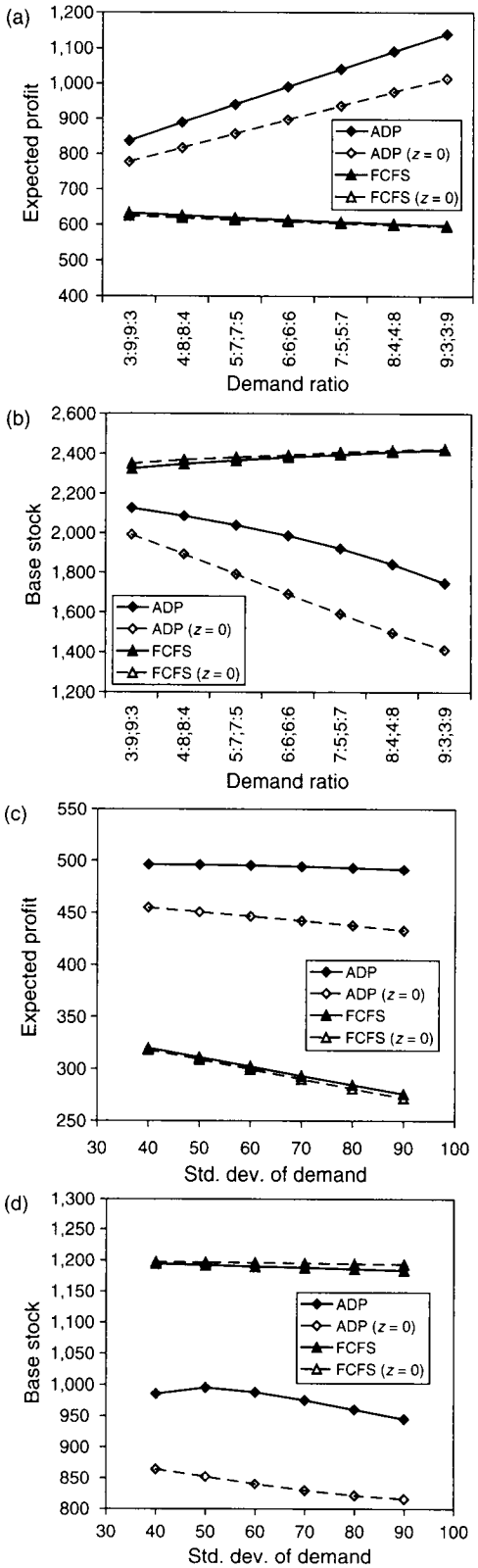
Table 2 compares the results of Example 1 for the ADP problem to the alternate cases. The main conclusions from the table are (1) the majority of the profit improvement over the naive solution, FCFS ($z = 0$), is attributable to the prioritization of the customer classes; (2) providing discounts in the ADP solution has a lesser, though significant, effect on profits; and (3) the discounts have a great effect on the class 1 fill rates, as they encourage the firm to increase its base-stock level and subsequently shift demand for class 1 customers from the first to the second stage.

In Figure 1 we display how changing the mean and variance of the demand affect the expected profit and base-stock levels in the solutions to Example 1. In Figures 1(a)

Table 2. Comparison of the ADP solution to the FCFS solution with and without price discounts (Example 1).

Measure	ADP	ADP ($z = 0$)	FCFS	FCFS ($z = 0$)
Expected profit	489.17	427.73	266.93	262.20
Base stock	932	813	1,182	1,192
Discount cost	17.02	0	2.73	0
Total fill rate %				
Average	99.98	65.80	97.08	94.46
Class 2	99.92	99.46	94.16	94.46
Class 1	100	32.16	100	94.46
Immediate fill rate %				
Average	71.59	65.80	93.94	94.46
Class 2	99.92	99.46	93.94	94.46
Class 1	43.26	32.16	93.94	94.46

Figure 1. (a) Expected profit vs. class 1 to class 2 demand ratio in the first and second stages; (b) Base-stock level vs. class 1 to class 2 demand ratio in the first and second stages; (c) Expected profit vs. demand variance; (d) Base-stock level vs. demand variance.



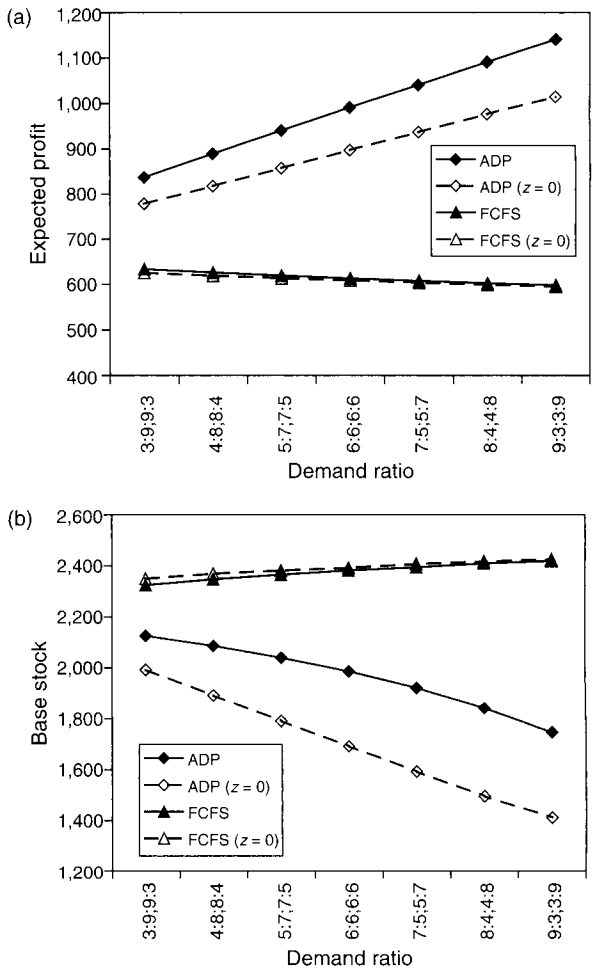
and (b), we change the ratio of the demand of class 1 to class 2 customers in the first and second stages. The ratios reported are $E[d_{11}]$:$E[d_{21}]$ and $E[d_{12}]$:$E[d_{22}]$. (Note that we let the mean total demand be 1,200 in each stage.) Observe that with higher stage 2, class 2 demand the expected profit in the ADP solution increases, but that of the FCFS solution decreases. In the ADP solution, the firm is able to reserve capacity for these customers. The higher profit results from avoiding the congestion cost in the first period, or the cost of price discounts needed to delay these customers to the second stage. In the FCFS solution, higher profits result when the class 2 customers are served in the first stage rather than the less valuable class 1 customers. Similarly, the base-stock level can be reduced with increased stage 2, class 2 demand in the ADP solution as more of the demand is revealed prior to having to serve it; the class 1 customers can be delayed if needed until the second stage. Observe that the reverse holds for the FCFS solution, as a higher base stock is needed to satisfy an increasing number of class 1 customers, while holding some inventory for class 2 customers arriving in the second stage.

In Figures 1(c) and (d), we observe that the expected profit of the ADP solution is robust to changes in the variability. Interestingly, the base-stock level for the ADP problem solution is not monotonic with the variance (as is the case with, say, a newsvendor model). With low amounts of variability, the firm can control which customer receives the marginal unit (the last unit allocated), so that a small increase in variability increases the base-stock level to ensure supply to class 2 customers. With greater increases in variability, the firm is increasingly unsure of which customer type would receive the marginal unit, lowering its value and thus lowering the base-stock level.

(Note that in these examples the base-stock level for FCFS is higher than that of the ADP solution; this need not be the case, and counterexamples can be constructed.)

EXAMPLE 2. In this example, we consider the scenario as in Example 1, but with $p = (10.2, 20)$, $c_p = 10$, and $h = 8$, and assume that demand d_{ij} is IID Poisson with mean $\lambda = 50$ for $i = 1, 2$ and $j = 1, 2$. Figures 2(a) and (b) summarize the desirable features of the ADP problem solution vis-à-vis the case without discounting and the FCFS case. We observe that the optimal base stock for the ADP problem is 19.7% lower than that of the FCFS solution. Further, the optimal profit for the ADP problem (at its optimal base-stock level) is 12.0% higher than the FCFS solution and 2.9% higher than the case with no discounts (at their respective optimal base-stock levels). The fill rate for the ADP solution (at its optimal base-stock level) is 19.7% higher than that of the ADP solution without discounting and 1.8% higher than that of the FCFS solution. Thus, while the expected profit depends to a greater degree on the reservation of inventory through the allocation mechanism, the total fill rate is mostly dependent on the provision of discounts to encourage waiting.

Figure 2. (a) Expected profit vs. class 1 to class 2 demand ratio in the first and second stages; (b) Base-stock level vs. class 1 to class 2 demand ratio in the first and second stages.



5. Heuristics

5.1. Expected Demand Heuristic

The computational time for solving the optimal ADP solution increases exponentially with the number of stages and customer classes. We therefore consider an expected demand heuristic (EDH) that assumes demand equals its expectation in each stage to quickly determine the price discount and allocation. The heuristic is as follows.

Expected Demand Heuristic (EDH). For stage j, denote the heuristic discount by z_{ij}^e for class $i \in I$ and the heuristic allocation Y_j^e. For stage M, z_{iM}^e can be found from Proposition 1(a) and $Y_M^e = \min[\bar{D}_M, X_M]$. For $j = 1, \ldots, M-1$, assume that demand equals its expectation, i.e., $d_{ij} = E[d_{ij}]$ for $i \in I$. Then, for each value of X_j and $\vec{D}_j$, we can determine through induction z_{ij}^e and Y_j^e using (8) and the inventory allocation rule. We can then find the base-stock level, X_1^e, under the EDH policy.

Table 3. EDH performance: Expected profits and base-stock levels.

Parameters		Expected profit			Base stock		
p_1	β_1	ADP	EDH	% Error	ADP	EDH	%Δ
1.2	5.0	457.14	457.14	0.00	955	955	0.00
1.4	1.25	504.93	495.74	1.82	913	908	0.54
1.6	0.3125	538.96	538.76	0.04	966	962	0.41
1.8	0.0781	616.91	616.90	0.00	1,175	1,177	0.17
2.0	0.0195	719.85	719.85	0.00	1,232	1,232	0.00

Because we only calculate the price discounts and the allocation policy once at each stage for each state X_j and $\vec{D}_j$, the computational time is reduced significantly. Moreover, because z_{ij}^e and Y_j^e are constant matrices with two dimensions, the system can make real-time decisions if we store their values.

To illustrate the accuracy of the EDH heuristic, we consider the case of Example 1 above, but let p_1 and $\beta_{11} = \beta_{12} = \beta$ vary, and present numerical results in Table 3. As can be seen, both the EDH expected profit and the initial inventory decision are very close (less than 2% deviation for the profit and less than 1% deviation for the initial inventory) to those of the optimal ADP solution. (The same results hold over larger numerical sets we experimented with, but are not reported here for the sake of brevity.)

5.2. A Continuous-Time Heuristic

In most practical cases, demand arrives according to a continuous-time stochastic process during a period and firms are required to respond to demands immediately, either accepting the order or offering a discount to encourage the customer to wait for delivery in the next period (or, of course, rejecting the demand). We could approximate such a case by dividing the period into a large number of stages, M, of say, equal duration. However, finding the solution as M increases is computationally difficult. Therefore, we consider a heuristic that determines whether a demand should be accepted based on the current inventory and the waiting demand under the assumption that all future demands will occur in a single stage. As we consider each demand separately, we must assume that there is no congestion cost, c_j.

For a continuous-time demand process, we consider a single arrival from class k at time t. Let $X(t)$ and $\vec{d}(t)$ denote the inventory level and the waiting demand at time t, and let $\alpha_k(t)$ and $\beta_k(t)$ denote the continuous-time analogs of α_{kj} and β_{kj}. We determine how to address the demand as follows.

Two-Stage Heuristic (TSH). We divide the period into two stages and consider the decision in the first stage. The demand in the first stage equals one unit of demand from class k, and the random demand in the second stage equals the r.v. for demand for the remaining time in the period. That is, we approximate the continuous-time problem by letting $d_{k1} = 1$, $d_{i1} = 0$ for $i \neq k$, $X_1 = X(t)$, $\vec{d}_1 = \vec{d}(t)$, and $d_{i2} = d_i(t, T]$ for $i \in I$ where $d_i(t, T]$ denotes the random variable of demand from class $i \in I$ customers in the time remaining in the period. Also, $\alpha_{k1} = \alpha_k(t)$, $\beta_{k1} = \beta_k(t)$, and $\alpha_{i2} = \alpha_i(T)$, $\beta_{i2} = \beta_i(T)$ for $i \in I$. We then determine if the customer demand should be accepted or, if not, we determine the discount to offer, z_{kj}^h. Note that because the congestion cost c_j equals 0, customers delayed immediate service are only allocated inventory at time T.

From Proposition 1(a) we can find the discounts, z_{i2}^h, to offer in the second stage as

$$z_{i2}^h = \left(\min\left[\frac{p_i + c_l - c_p - c_w}{2} - \frac{\alpha_{i2}}{2\beta_{i2}}, \frac{1-\alpha_{i2}}{\beta_{i2}}\right]\right)^+$$

for each $i \in I$, and from Proposition 3 we can find

$$z_{k1}^h = \left(\min\left[\frac{p_k + c_l - c_p - c_w}{2} - \frac{\alpha_{k1}}{2\beta_{k1}} + \frac{1}{2}(h + c_w) \cdot P\left(\sum_{i\in I} d_{i2} + \vec{d}_1 < X_1\right), \frac{1-\alpha_{k1}}{\beta_{k1}}\right]\right)^+.$$

From the inventory allocation rule, letting $\gamma_{k1} = \alpha_{k1} + \beta_{k1} z_{k1}^h$, let X_k^h be the value of X that solves

$$p_k + c_l - \gamma_{k1}\left(p_k + c_l - z_{k1}^h - c_p - c_w + (h + c_w) \cdot P\left(\sum_{i\in I} d_{i2} + \vec{d}_1 < X\right)\right)$$

$$= E_{\vec{d}_2}[\Delta_X \Pi_2(X, \vec{d}_2)]$$

$$= \sum_{i\in I} L_{i2} P(X \in [D_{i+1,2}, D_{i2})) + (c_p + c_w) P(X \in [D_{12}, \bar{D}_2)) + (c_p - h) P(X \geqslant \bar{D}_2).$$

We allocate a unit of $X(t)$ to the class k customer if and only if $X(t) \geqslant X_k^h$. By Proposition 2, X_k^h is the minimum amount of inventory needed for the future demand.

The base-stock level for the TSH is given by solving a single-stage problem assuming that demand for each class in the stage is distributed according to the demand for the entire period, i.e., letting $d_{i1} = d_i[0, T]$. □

The TSH requires solving a single-stage problem representing the future demand for any time t. Therefore, it can be applied in real time. Further, because the values of $z_i^h(t)$ and $X_i^h(t)$ can be stored for bounded demand functions and a finite, but large, number of times t, a simple table lookup procedure can be implemented to give a very close approximation to the TSH.

EXAMPLE 3. To illustrate the effectiveness of the TSH heuristic, we present computational results in Table 4. We let $K = 2$, $c_p = 10$, $h = 8$, $c_w = c_l = 0$, $p_2 = 20$, and $\alpha = (0, 0)$ and $\beta_2 = 0.01$. We vary the parameters p_1

Table 4. Comparison of the expected profit and total fill rate for the ADP solution and the two-stage heuristic.

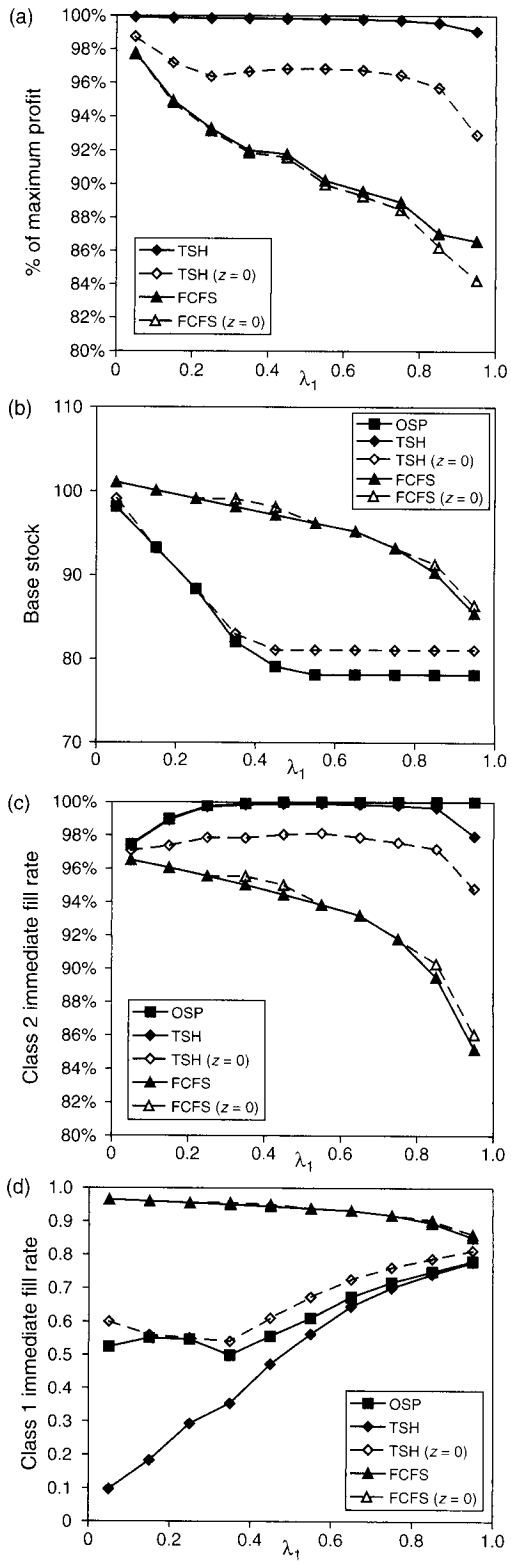
Parameters		Expected profit		Total fill rate %			
				Class 1		Class 2	
p	β	ADP	TSH	ADP	TSH	ADP	TSH
12	5.0000	274.62	273.55	100	100	100	99.07
14	1.2500	446.48	444.60	100	100	100	98.47
16	0.3125	603.18	601.13	99.52	99.66	100	98.11
18	0.0781	758.85	757.44	96.99	97.24	100	98.19
20	0.0195	926.62	926.16	96.98	97.19	100	98.08

and β_1. Demand $d_1(t)$ and $d_2(t)$ are independent Poisson processes with arrival rates $\lambda_1 = 0.9$ and $\lambda_2 = 0.1$. We consider a period of length $T = 100$. We compare the results of the TSH to an upper bound for the optimal profit of the ADP given by the solution to a one-stage problem (OSP), where all demand information is known prior to allocating inventory and determining discounts. From Table 4, it can be seen that the TSH was within 0.5% of the upper bound. Further, the TSH solution resulted in total fill rates per class (i.e., total demand per class eventually met) that were within 2% of those provided by the optimal ADP solution, with the TSH solution having higher fill rates for class 1 and lower for class 2 than in the optimal solution. (Similar results were observed for other cases but not reported, again for the sake of brevity.)

We next compare the results of the TSH to the FCFS solution and the OSP upper bound. Using the parameter values in Example 3 and letting $p = (10.2, 20)$ and $\beta = (10.0, 0.01)$, we vary the demand arrival rates, λ_1 and λ_2, keeping their sum equal to 1.

Figures 3(a)–(d) show the change in the expected profit as a percentage of the single-stage problem upper bound, the base-stock level, and the fill rates as $\lambda_1 = 1 - \lambda_2$ increases. The figures show that the TSH performs very well compared to the upper bound and confirms our previous result that the majority of the profit improvement over the naive solution is attributable to customer prioritization. This profit is achieved with a lower base-stock level then the FCFS solution. The base stock for the TSH (which is coincident with that of the OSP) decreases as the relative level of class 1 demand increases because the firm is able to reserve capacity for the class 2 demand. The base stock then levels off at an appropriate value when the last customer served immediately is most likely to be a class 1 customer. We observe in Figure 3(c) that the class 2 immediate fill rate in the TSH is very high, with only some decline in cases when class 2 customers represent a clear majority or minority, i.e., when there are either too many class 2 customers to serve all or too few to worry about. Of course, this advantage comes at the price of immediate service to class 1 customers (Figure 3(d)).

Figure 3. Varying the class 1 demand arrival rate, λ_1 $(=1 - \lambda_2)$, showing (a) Expected profit as a percentage of the one-stage-problem (OSP) upper bound; (b) Base-stock level; (c) Class 2 immediate service rate; and (d) Class 1 immediate service rate.



6. Conclusions

The model in this paper considers the partial-backlogging case for a multiple-customer-class inventory system where the likelihood of backlogging is a linear function of an offered discount. This partial backlogging creates dependencies in the stages so that a supplier must consider both inventory position and current waiting customers prior to determining the inventory allocation. Previous research (Topkis 1968) has considered the lost-sales case of the problem. The paper contributes to the literature by developing algorithms to determine both the optimal discounts to offer and the quantity to allocate in each period. By dynamically adjusting the discount offered, we show that the supplier can effectively address the problem.

The problem is of interest in that it captures a fundamental trade-off suppliers must make between serving demand at hand and potentially more lucrative future demand. Such problems are seen in product distribution center operations where there is a need to commit to fulfillment of demand. However, in practice, customers may be willing to take delayed delivery in exchange for a price reduction. Similarly, in practice a supplier must be able to address requests on a real-time basis instead of batching decisions. We show that a two-stage heuristic that addresses individual decisions based on the current inventory and waiting customer state and expectations for the remainder of the period performs very well.

Several assumptions are made in the solution of the problem, some of which can be relaxed with only minor changes in the solution, while others may require greater effort. Among the former, we assume that the cost of lost sales and cost of delaying demand are constant in every stage. However, a customer's interpretation of denial of service early in a period would differ from being refused service at the end of the period, where explanations of a stockout might be reasonable. Making these costs time dependent would require only minor changes in the model. Also, we assume unit demands. However, as long as the number of customers is much larger than the total inventory to allocate, relaxing this assumption should not change the results, and it would make the analysis much more difficult. On the other hand, we have made the assumption that customers whose contracted price is high are less willing to wait if offered a discount. While such an assumption is reasonable in a contracting setting where prices are set to reflect expected service levels, in a noncontracted price environment other assumptions might be more reasonable. For example, if customers are searching for product and express willingness to pay higher prices, they may also be more willing to wait for delivery (perhaps as the supplier is the only distributor identified). Future research may address alternate assumptions.

Appendix

PROOF OF PROPOSITION 1. (a) Given Y_M and ω_M (and system state X_M and $\bar{d}_M$), the optimal discounts satisfy (7a) and (7b). The first-order conditions on the unconstrained objective (7a) are

$$\begin{aligned}
&\frac{\partial \pi_M}{\partial z_{iM}} \\
&= \sum_{l=1}^{d_{iM}} \frac{\partial \pi_{liM}}{\partial z_{iM}} \\
&= \sum_{l \text{ s.t. } Y_M < D_{liM}} \frac{\partial((\alpha_{iM}+\beta_{iM}z_{iM})(p_i - z_{iM} - c_p - c_w) - (1-\alpha_{iM}-\beta_{iM}z_{iM})c_l)}{\partial z_{iM}} \\
&= \sum_{l \text{ s.t. } Y_M < D_{liM}} \beta_{iM}(p_i + c_l - c_p - c_w) - \alpha_{iM} - 2\beta_{iM}z_{iM} \\
&= 0
\end{aligned}$$

for all $i = 1, \ldots, K$ that give z'_{iM}. Noting the concavity, constraining the solution by (7b) provides the solution.

(b) The proof is similar to the proof of Proposition 4 for the general stage j, and we delay the presentation until then.

(c) Allocating the inventory to all of the demand for the last stage in the period follows directly from the assumption $c_w + h \geqslant c_M$. □

PROOF OF PROPOSITION 2. (a) Let $\Delta_X \pi_{liM}$ be the change in the profit contributed by the lth customer of class i ($i = 0, \ldots, K$) in stage M when an additional unit of inventory is available. Because at most one customer would receive the unit, at most one such term is nonzero. Further, from Proposition 1, $Y_M^* = \min[X_M, \bar{D}_M]$. Therefore, from the definitions for L_{ij} and letting $D_{K+1,M} = 0$ and noting $\bar{D}_M = D_{0,M}$,

$$\begin{aligned}
\Delta_X \Pi_M &= \sum_{i=0}^{K} \left(\sum_{l=1}^{d_{iM}} \Delta_X \pi_{liM} \right) + (c_p - h) \cdot 1_{X_M \geqslant \bar{D}_M} \\
&= \begin{cases} L_{iM}(z_{iM}^*, \omega_M^*, Y_M^* \mid X_M, \bar{d}_M) \\ \qquad \text{if } D_{i+1,M} \leqslant X_M < D_{iM} \text{ for } i = 0, \ldots, K, \\ c_p - h, \quad \bar{D}_M \leqslant X_M. \end{cases}
\end{aligned}$$

Observe that

(i) $c_p - h \leqslant L_{0,M} = c_p + c_w - c_M \cdot 1_{X_M \geqslant g_M}$ because $c_M \leqslant c_w + h$;

(ii)

$$\begin{aligned}
L_{0,M} &= c_p + c_w - c_M \cdot 1_{X_M \geqslant g_M} \leqslant L_{1,M}(z_{iM}^*, \omega_M^*, Y_M^* \mid X_M, \bar{d}_M) \\
&= p_1 + c_l - \gamma_{1,M}^*(p_1 + c_l - z_{1,M}^* - c_p - c_w) - c_M \cdot 1_{X_M \geqslant g_M}
\end{aligned}$$

because this implies $(1 - \gamma_{1,M}^*)(c_p + c_w) \leqslant (1 - \gamma_{1M}^*)(p_1 + c_l) + \gamma_{1M}^* z_{1,M}^*$, which follows from the assumption $p_1 + c_l - c_p - c_w > 0$; and finally,

(iii) $L_{i+1,M} \geqslant L_{iM}$ by Proposition 1. Therefore, $\Delta_X \Pi(X_M, \bar{d}_M)$ is nonincreasing in X_M and $\Delta_X \Pi(X_M, \bar{d}_M)$ is nondecreasing in $\bar{D}_M$.

(b) Designating the demand in stage M as $\bar{d}_M = \{d_M, d_{0,M}\}$,

$$\Delta_{\bar{D}}\Pi_M = \Pi_M(X, \{d_M, d_{0,M}+1\}) - \Pi_M(X, \{d_M, d_{0,M}\}) = \begin{cases} c_w + h - c_M \cdot 1_{\bar{D}_M \geqslant g_M}, & X_M > \bar{D}_M, \\ 0, & X_M \leqslant \bar{D}_M. \end{cases}$$

Then,

$$E_{\bar{d}}[\Delta_{\bar{D}}\Pi_M] = (c_w + h)P(\bar{D}_M < X_M) - c_M P(g_M \leqslant \bar{D}_M < X_M) \quad (11)$$

$$\geqslant (c_w + h - c_M)P(\bar{D}_M < X_M)$$

$$\geqslant 0 \quad (12)$$

and $E_{\bar{d}}[\Delta_{\bar{D}}\Pi_M] \leqslant c_p + c_w$ because $c_p > h$. From (11), observe that $E_{\bar{d}}[\Delta_{\bar{D}}\Pi_M]$ is nondecreasing in X_M and nonincreasing in $\bar{D}_M$. □

Proof of Proposition 4. Consider a priority sequence ω where the lth customer of class i is followed by the mth customer of class k, where $L_{kj} > L_{ij}$ and $Y = D^{\omega}_{lij}$. Then, by interchanging the two customers, the cost can be reduced. Similarly, if the customers are arranged in nonincreasing order of the cost of delaying them (i.e., $D^{\omega}_{lij} \leqslant D^{\omega}_{mkj}$ if $L_{ij} \geqslant L_{kj}$ for all i, k, l, m.), then by choosing any set of customers and rearranging them, the firm cannot reduce the cost. Thus, a priority sequence is optimal for all allocations Y_j if and only if the values L_{ij} are nonincreasing with the order.

Let $x, y \in I$ be two customer classes with $x > y$. It remains to be shown that $L_{xj} \geqslant L_{yj}$ and $L_{xj} \geqslant L_{0,j}$ for all x and y. Let $\gamma_{yj}(z) = \alpha_y + \beta_y z$ for any z such that $0 \leqslant z \leqslant (1-\alpha_y)/\beta_y$. We know that z'_{yj} maximizes $\gamma_{yj}(z)(p_i + c_l - z - c_p - c_w + E[\Delta_{\bar{D}}\Pi])$ and that z^*_{yj} equals either 0, z'_{yj}, or $(1-\alpha_{yj})/\beta_{yj}$. Therefore, for any z_y,

$$L_{xj} - L_{yj} \geqslant (p_x + c_l - \gamma^*_{xj}(p_x + c_l - z^*_{xj} - c_p - c_w + E[\Delta_{\bar{D}}\Pi(X_j - Y_j, \bar{d}_{j+1})])) - (p_y + c_l - \gamma_{yj}(z_y)(p_y + c_l - z_y - c_p - c_w + E[\Delta_{\bar{D}}\Pi(X_j - Y_j, \bar{d}_{j+1})])).$$

Letting $z_y = \min[z^*_{xj}, (1-\alpha_y)/\beta_y]$ implies $\gamma''_{yj} \equiv \gamma_{yj}(z_y) = \min[\alpha_y + \beta_y z^*_{xj}, 1] \geqslant \alpha_x + \beta_x z^*_{xj} = \gamma^*_{xj}$ because $\alpha_y \geqslant \alpha_x$ and $\beta_y \geqslant \beta_x$. Therefore,

$$L_x - L_y \geqslant (1-\gamma''_{yj})(p_x - p_y) + \gamma''_{yj}(z^*_{xj} - z_y) + (\gamma''_{yj} - \gamma^*_{xj}) \cdot (p_x + c_l - z^*_{xj} - c_p - c_w + E[\Delta_{\bar{D}}\Pi(X_j - Y_j, \bar{d}_{j+1})]) \geqslant 0$$

because $z^*_{xj} \leqslant (p_x + c_l - c_p - c_w + E[\Delta_{\bar{D}}\Pi(X_j - Y_j, \bar{d}_{j+1})])^+$ by Proposition 3, $p_x + c_l - c_p - c_w \geqslant 0$ by assumption, and $E[\Delta_{\bar{D}}\Pi(X_j - Y_j, \bar{d}_{j+1})] \geqslant 0$ by the induction assumption. Also,

$$L_{xj} - L_{0,j} = (1-\gamma^*_{xj})(p_x + c_l - c_p - c_w + E[\Delta_{\bar{D}}\Pi(X_j - Y_j, \bar{d}_{j+1})]) + \gamma^*_{xj} z^*_{xj} \geqslant 0$$

by similar reasoning. □

Proof of Proposition 6. (a) If $X_j < \bar{D}_j$, an incremental unit of inventory allocated to stage j results in

$$\Delta_X\Pi_j(z^*_{ij}, \omega^*_j, Y^*_j + 1 \mid X_j + 1, \bar{d}_j) = \sum_{i=0}^{K} L_{ij} \cdot 1_{D_{i+1,j} \leqslant Y^*_j < D_{ij}} - c_j \cdot 1_{g_j \leqslant Y^*_j < \bar{D}_j}.$$

(Note that from the notational definition, $\bar{D}_j = D_{0,j}$.) If the incremental unit of inventory is allocated to stage $j+1$, then

$$\Delta_X\Pi_j = E_{\bar{d}_{j+1}}[\Delta_X\Pi_{j+1}].$$

Because the unit should be allocated to the stage with the higher incremental profit, then from the induction assumption

$$\Delta_X\Pi_j = \max\left[\sum_{i=0}^{K} L_{ij} \cdot 1_{D_{i+1,j} \leqslant Y^*_j < D_{ij}} - c_j \cdot 1_{g_j \leqslant Y^*_j < \bar{D}_j}, E_{\bar{d}_{j+1}}[\Delta_X\Pi_{j+1}]\right] \geqslant E_{\bar{d}_{j+1}}[\Delta_X\Pi_{j+1}] \geqslant c_p - h \geqslant 0.$$

Because L_{ij} and $E_{\bar{d}_{j+1}}[\Delta_X\Pi_{j+1}]$ are nonincreasing in X_j and nondecreasing in $\bar{D}_j$, then both are preserved under expectation, and so (a) holds.

(b) A similar approach gives

$$\Delta_{\bar{D}}\Pi_j(X_j, \bar{d}_j) = \max[c_p + c_w - E_{\bar{d}_{j+1}}[\Delta_X\Pi_{j+1}] - c_j \cdot 1_{g_j \leqslant \bar{D}_j}, E_{\bar{d}_{j+1}}[\Delta_{\bar{D}}\Pi_{j+1}]],$$

and the second result holds by the same reasoning as the first. □

Acknowledgments

The authors thank the editors and referees for their timely review and for many helpful comments that greatly improved the paper. Panos Kouvelis acknowledges support from the Boeing Center on Technology, Information and Manufacturing at the Olin School of Business for this research project. Joseph Milner was supported by a grant from the National Sciences and Engineering Research Council (NSERC) of Canada.

References

Alderman, H. 1987. Allocation of goods through non-price mechanisms: Evidence on distribution by willingness to wait. *J. Development Econom.* **25** 105–124.

Barzel, Y. 1974. A theory of rationing by waiting. *J. Law Econom.* **18** 73–95.

Belobaba, P. P. 1987. Airline yield management: An overview of seat inventory control. *Transportation Sci.* **21** 63–73.

Bitran, G. R., S. V. Mondschein. 1997. Periodic pricing of seasonal products in retailing. *Management Sci.* **43** 64–79.

Bruce, N., P. Desai, R. Staelin. 2004. Enabling the willing: Consumer rebates for durable goods. *Marketing Sci.* Forthcoming.

Brumelle, S. L., J. I. McGill. 1993. Airline seat allocation with multiple nested fare classes. *Oper. Res.* **41** 127–137.

Cattani, K. D., G. C. Souza. 2002. Inventory rationing and shipment flexibility alternatives for direct market firms. *Production Oper. Management* **11**(4) 441–457.

Chan, L. M. A., D. Simchi-Levi, J. Swann. 2005. Dynamic pricing strategies for manufacturing with stochastic demand and discretionary sales. *Manufacturing Service Oper. Management.* Forthcoming.

Chen, F. 2001. Market segmentation, advanced demand information, and supply chain performance. *Manufacturing Service Oper. Management* **3**(1) 53–67.

Cheung, K. L. 1998. A continuous review inventory model with a time discount. *IIE Trans.* **30** 747–757.

Cohen, M. A., P. R. Kleindorfer, H. L. Lee. 1988. Service constrained (s, S) inventory systems with priority demand classes and lost sales. *Management Sci.* **34** 482–499.

Dana Jr., J. D. 1999. Using yield management to shift demand when the peak time is unknown. *RAND J. Econom.* **30**(3) 456–474.

Deacon, R. T., J. Sonstelie. 1985. Rationing by waiting and the value of time: Results from a natural experiment. *J. Political Econom.* **93** 627–647.

DeCroix, G. A., A. Arreola-Risa. 1998. On offering economic incentives to backorder. *IIE Trans.* **30** 715–721.

Deshpande, G. V., M. A. Cohen, K. Donohue. 2003. A threshold inventory rationing policy for service-differentiated demand classes. *Management Sci.* **49**(6) 683–703.

de Vericourt, F., F. Karaesmen, Y. Dallery. 2001. Assessing the benefits of different stock-allocation policies for a make-to-stock production system. *Manufacturing Service Oper. Management* **3**(1) 53–67.

Federgruen, A., A. Heching. 1999. Combined pricing and inventory control under uncertainty. *Oper. Res.* **47** 454–475.

Feng, Y., B. Xiao. 2000. Optimal policies of yield management with multiple predetermined prices. *Oper. Res.* **48** 332–343.

Frank, K. C., R. Q. Zhang, I. Duenyas. 2003. Optimal policies for inventory systems with priority demand classes. *Oper. Res.* **51**(6) 993–1002.

Gale, I. L., T. J. Holmes. 1993. Advance-purchase discounts and monopoly allocation of capacity. *Amer. Econom. Rev.* **83**(1) 135–146.

Gallego, G., G. J. van Ryzin. 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizon. *Management Sci.* **40** 999–1020.

Gerchak, Y., M. Parlar, T. K. M. Yee. 1985. Optimal rationing policies and production quantities for products with several demand classes. *Canadian J. Admin. Sci.* **2** 161–176.

Ha, A. 1997. Inventory rationing in a make-to-stock production system with several demand classes and lost sales. *Management Sci.* **43** 1093–1103.

McGill, J. I., G. J. van Ryzin. 1999. Revenue management: Research overview and prospects. *Transportation Sci.* **33** 233–256.

Png, I. P. L., D. Reitman. 1994. Service time competition. *RAND J. Econom.* **79**(2) 619–634.

Robinson, L. W. 1995. Optimal and approximate control policies for airline booking with sequential nonmonotonic fare classes. *Oper. Res.* **43** 252–263.

Thowsen, G. T. 1975. A dynamic, nonstationary inventory problem for a price/quantity setting firm. *Naval Res. Logist.* **22** 461–476.

Topkis, D. M. 1968. Optimal ordering and rationing policies in a nonstationary dynamic inventory model with n demand classes. *Management Sci.* **15** 160–176.

Wang, Y., M. A. Cohen, Y. S. Zheng. 2002. Differentiating customer service on the basis of delivery lead times. *IIE Trans.* **34**(11) 979–989.

Weatherford, L. R., S. Bodily, P. Pfeiffer. 1993. Modeling the customer arrival process and comparing decision rules in perishable asset revenue management situations. *Transportation Sci.* **27** 239–251.

Wollmer, R. D. 1992. An airline seat management model for a single leg route when lower fare classes book first. *Oper. Res.* **40** 26–37.

Zabel, E. 1972. Multi-period monopoly under uncertainty. *J. Econom. Theory* **5** 524–536.

Zhao, W., Y. S. Zheng. 2000. Optimal dynamic pricing of perishable assets with nonhomogeneous demand. *Management Sci.* **46** 375–388.

OPERATIONS RESEARCH
Vol. 54, No. 1, January–February 2006, pp. 198–200
ISSN 0030-364X | EISSN 1526-5463 | 06 | 5401 | 0198

informs

DOI 10.1287/opre.1060.0276

Contributors

Belarmino Adenso-Díaz ("Fine-Tuning of Algorithms Using Fractional Experimental Designs and Local Search") is a Professor in the Engineering School, University of Oviedo, Spain. His experience is that fine-tuning algorithms is one of the most tedious tasks in the development and implementation of optimization procedures. The idea for this paper surfaced in conversations about how to use exiting operations research tools to facilitate building additional ones.

F. Babonneau ("Solving Large-Scale Linear Multicommodity Flow Problems with an Active Set Strategy and Proximal-ACCPM") is a Ph.D. student at the University of Geneva. He is interested in developing solution methods for large-scale convex optimization problems. The present work is part of his Ph.D. thesis, which is devoted to solving linear and nonlinear multicommodity flow problems using the analytic center cutting plane method.

Erin Baker ("Increasing Risk and Increasing Informativeness: Equivalence Theorems") is an Assistant Professor in the Department of Mechanical and Industrial Engineering at University of Massachusetts, Amherst. She is interested in decision making under uncertainty, with applications to climate change policy and technology R&D. This research was part of her Ph.D. thesis, conducted at Stanford University under the supervision of John Weyant, James Sweeney, Susan Athey, and Jonathon Levin.

Dimitris Bertsimas ("A Robust Optimization Approach to Inventory Theory") is the Boeing Professor of Operations Research at the Sloan School of Management and the Operations Research Center at the Massachusetts Institute of Technology. His current research focuses on robust optimization as a tractable theory for optimization under uncertainty.

Qing Ding ("Dynamic Pricing Through Discounts for Optimizing Multiple-Class Demand Fulfillment") is an Assistant Professor of Operations Management at the School of Business, Singapore Management University. The paper is part of a dissertation written under the direction of Panos Kouvelis and Joseph Milner.

O. du Merle ("Solving Large-Scale Linear Multicommodity Flow Problems with an Active Set Strategy and Proximal-ACCPM") is the Operational Research Director of Air France. He holds a Ph.D. in operations research from the University of Geneva. He has contributed to new methods for column generation schemes and is interested in applying them to airline problems.

Talat Genc ("A Stochastic Programming Approach to Power Portfolio Optimization") is a Professor at the Department of Economics, University of Guelph, Canada. Professor Genc's research interests are in energy economics, industrial organization, and econometrics, especially in the context of problems arising in the electric power industry. He is also interested in modeling uncertainty within game theoretic models. The statistical and econometric aspects of this paper were developed as parts of his dissertation at the University of Arizona.

Peter W. Glynn ("A Nonparametric Approach to Multiproduct Pricing") is the Thomas Ford Professor of Engineering in the Department of Management Science and Engineering at Stanford University and has a courtesy appointment in the Department of Electrical Engineering. He is a Fellow of the Institute of Mathematical Statistics. His research interests include computational probability, queuing theory, statistical inference for stochastic processes, and stochastic modeling.

Linda V. Green ("Managing Patient Service in a Diagnostic Medical Facility") is the Annand G. Erpf Professor of Business at the Graduate School of Business, Columbia University. This work is a part of a series of papers focusing on applications of operations research in the healthcare industry. Her current research focuses on improving emergency responsiveness and identifying strategies for the effective design and management of diagnostic facilities.

Ahmed Hadjar ("A Branch-and-Cut Algorithm for the Multiple Depot Vehicle Scheduling Problem") is a researcher at the École Polytechnique de Montréal and the Groupe d'Études et de Recherche en Analyse des Décisions (GERAD). He received a Ph.D. in operations research from the Institut National Polytechnique de Grenoble (France). His research interests are in combinatorial optimization and integer programming, especially applications to routing and scheduling.

L. Jeff Hong ("Discrete Optimization via Simulation Using COMPASS") is an Assistant Professor in the Department of Industrial Engineering and Logistics Management at the Hong Kong University of Science and Technology. His research interests include optimization via simulation, simulation experimental design, and simulation output analysis. This paper is based on a chapter in his Ph.D. dissertation.

Panos Kouvelis ("Dynamic Pricing Through Discounts for Optimizing Multiple-Class Demand Fulfillment") is the

Operations Research 54(1), pp. 198–200, © 2006 INFORMS

Aurélie Thiele ("A Robust Optimization Approach to Inventory Theory") is an Assistant Professor in the Department of Industrial and Systems Engineering at Lehigh University. Her research focuses on the development and analysis of tractable models of uncertainty for dynamic optimization problems in operations management. This work was part of her Ph.D. dissertation at the Massachusetts Institute of Technology under the supervision of Dimitris Bertsimas.

Benjamin Van Roy ("A Nonparametric Approach to Multiproduct Pricing") is an Assistant Professor of Management Science and Engineering, Electrical Engineering, and by courtesy, Computer Science at Stanford University. His recent research interests include stochastic control, machine learning, economics, finance, and information technology.

J.-P. Vial ("Solving Large-Scale Linear Multicommodity Flow Problems with an Active Set Strategy and Proximal-ACCPM") is a Professor of Operations Management at the University of Geneva. He is interested in algorithms for convex optimization and their applications to logistics.

Ben Wang ("Managing Patient Service in a Diagnostic Medical Facility") holds a Ph.D. in Decision, Risk, and Operations from Columbia Business School. He previously worked as an associate at the Industrial & Commercial Bank of China and is currently with Century Securities of China.

Ward Whitt ("Fluid Models for Multiserver Queues with Abandonments") is a Professor at Columbia University in the Department of Industrial Engineering and Operations Research. He joined the faculty of Columbia University after spending 25 years in research at AT&T.

Lihua Yu ("A Stochastic Programming Approach to Power Portfolio Optimization") is a Quantitative Analyst at Pennsylvania Power and Light (PPL). His research interests are in applications of operations research and related quantitative tools to problems arising in the electric power industry. Prior to joining PPL, he was a research associate at the State Utility Forecasting Group at Purdue University. The optimization algorithm reported in this paper was part of his dissertation research at the University of Arizona.