On the time-window fulfillment rate in a single-item min-max inventory control system

TONG WANG¹, YOUHUA CHEN², and YOUYI FENG²

¹INSEAD, Fontainebleau 77305, France
E-mail: tong.wang@insead.edu

²Department of System Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, NT, Hong Kong
E-mail: {yhcchen or yyfeng}@se.cuhk.edu.hk

Received January 2003 and accepted June 2004

This paper considers an inventory control system with a constraint on the service level. We focus on a popular service target in practice, the time-window ready rate, which is defined as the percentage of periods in which demands are completely fulfilled within a pre-specified time window. We explore the cost impact of different time-window ready rate settings by proposing evaluation and optimization methods for a single-location single-item min-max (or \((s, S)\)) inventory system. New optimization procedures have been developed because the existing methods, which are based on a Lagrangian multiplier, cannot generally handle the current problem. More specifically, in some cases the time-window ready rate is a complicated function of the re-order point \(s\), which renders existing algorithms inapplicable. Our algorithms are based on the monotone properties of cost and service constraint functions, which allow us to search the optimal policy efficiently within an identified region of policy-parameter values. Numerical experiments show the non-linear trade-off between cost and service, and indicate that extending or shrinking the time window can have a significant impact on the inventory cost.

1. Introduction

Timely order fulfillment is a ubiquitous customer service criterion in the manufacturing and distribution sectors. This criterion is commonly measured by the time-window fulfillment rate, which is defined as the percentage of times that orders are completely fulfilled within a specific time window. Many companies have set targets for their order fulfillment. For example, around 1995, Hewlett-Packard targeted a 93% fulfillment rate within 3 days, and IBM PC and Compaq aimed at accomplishing 95% within 5 days (Hausman et al. 1998).

In the academic arena, many operations management textbooks discuss two types of fulfillment or service levels: fill rate and ready rate (Nahmias, 2001). Fill rate is commonly defined as the fraction of quantity fulfilled over quantity requested, whereas ready rate is frequency based, measuring the percentage of times in which the system can fulfill all demands on time (i.e., instantly/off-shelf or within a specified time window). Service levels are also measured on both an item and order basis. An order-based service level focuses on the fulfillment of a whole order that may be comprised of multiple items, whereas an item-based service level looks only at the fulfillment of a particular item. The dominant order-based service level is the order fill rate which corresponds to the probability of joint demand fulfillment within a given time window. Because the quantities of different items cannot be aggregated, whether an order is fulfilled or not depends on whether all the requested items are completely met. Thus, the item-based counterpart of order fill rate is actually the ready rate.

It is well known that controlling service levels for individual items is computationally convenient, and that the order-based fill rate can be approximated as a weighted average of item ready rates (Ballou, 1999). For instance, consider a two-item inventory system from which customers order either one of the two items or both. Statistical analysis indicates that the frequency of ordering each of the two items is 0.2, whereas that of ordering both items is 0.6. If the ready rates of individual items (within a time window) are all 0.95, then the weighted order fill rate is 0.95 × 0.4 + 0.95 × 0.95 × 0.6 = 0.92. If the target weighted fill rate is specified, then the service levels for each item must be adjusted to achieve the desired order fill rate. In a continuous review setting with Poisson demand arrivals, (Song, 1998) shows analytically that such a weighted average constitutes a lower bound on the exact order fill rate. In an intensive numerical study, she further shows that the bound provides a good approximation. For this reason, we address the item-based ready rate in this paper.

There are essentially two approaches to the modelling of inventory costs. The first is the so-called full cost approach
which assumes setup (replenishment), holding and shortage costs. The second is the so-called service-level approach which considers only setup and holding costs while taking one or more service-level constraints as surrogates for shortage costs. The use of service-level constraints is mainly motivated by the difficulty of estimating shortage costs, for which the service-level concept is viewed as an appealing practical alternative. Although specifying service-levels can be equally challenging, the service-level approach is widely adopted (Lee and Billington, 1992, 1995; Taylor, 1997). This paper also takes the service-level approach.

The objective of this paper is twofold. First, we address the evaluation of the time-window ready rate. Second, we provide procedures for optimizing the inventory control policy in a single-location, single-item inventory model with an item-based time-window ready rate constraint. In setting the target for the time-window ready rate, managers typically need to know the trade-off between service level and total inventory cost, where the total cost is the sum of the holding and ordering costs. They try to strike a balance between cost and service so that both the financial viability of the operation and its competitiveness in the market place are well maintained. As a time-window ready rate essentially involves two parameters, one for the time window and the other for the ready rate, it is useful to evaluate the costs of different time-span ready-rate combinations. Furthermore, in understanding the trade-off between cost and service, managers are able to recommend how much cost needs to be absorbed (or can be passed over to customers by a discount) if shortening (or extending) the delivery time window.

The time window $T$ ready rate is often equivalent to the off-shelf ready rate in a system in which the leadtime, $L$, is truncated by the time window, $T$. (See the work of Sherbrooke (1992) and Kruse (1981) for single-item systems, and Hausman et al. (1998) and Song (1998) for multi-item systems.) However, this equivalence holds only for systems with base-stock or order-up-to policies. For example, if the replenishment leadtime is shorter than the time window of fulfillment ($L < T$), then the equivalence is unwarranted. At first glance, this would seem trivial. However, it is actually problematic. When a re-order point, order-up-to level policy is adopted, the equivalence holds true only when the re-order point is non-negative. Specifically, when the reorder level $s$ is below zero, the expression for the ready rate is intricate, and the previously developed evaluation methods do not work properly in such a context. As the $L < T$ case may arise in multi-item systems, it should be considered. Consider an example of a two-item system: the leadtime for item $A$ is 1 week, and that for item $B$ is 4 weeks. If the firm sets a fulfillment window of 2 weeks, then the above scenario arises in item $A$.

A service-level-constrained inventory model is typically handled by a Lagrangian multiplier: the shadow cost of service constraint. In our problem, the shadow cost is a penalty to be “paid” if demand is not completely met within the target time-window. Combining holding cost and the shadow cost that is associated with the service constraint forms the imputed single-period loss function. When the imputed loss function is convex (or quasi-convex), many optimization procedures are available for obtaining an optimal solution that minimizes the total cost and satisfies the service-level constraint. For the service level that we propose here, however, the imputed loss function may fail to be convex (or quasi-convex). In this paper we apply an exact direct method for the model optimization, one that can obtain the optimal policy parameter values through a guided search over an identified two-dimensional space.

At the end of this paper we conduct a numerical study which contains several thousands of examples. It shows that a significant amount of cost, mostly more than 10%, can be saved if the time window can be expanded by one period. When the market is competitive, it is most likely that a firm will need to set a high service level within its promised fulfillment window. Since there is a significant trade-off to be made between the window length and inventory costs, the firm might consider offering deep discounts to those price-sensitive but time-insensitive customers in exchange for longer fulfillment windows. This analysis provides useful tools for the design and control of such systems.

The rest of the paper is organized as follows. Section 2 provides a literature review, and the model and mathematical expressions of average cost and service levels are described in Section 3. The properties of the $(s, S)$ policies are revealed and the algorithms are developed in Section 4.
Numerical examples are summarized and reported in Section 5. The last section discusses several possible extensions.

2. Literature review

The cost-service trade-off or exchange curve lies at the center of inventory management. Many published studies have addressed the trade-off analysis. Broadly speaking, service levels can be classified into two categories: off-shelf and time window, whereas off-shelf-based service levels are popular in textbooks, their time-window counterparts are the most common in practice (Lee and Billington, 1992; Taylor, 1997). Furthermore, as mentioned earlier, service levels can be defined at both the item and order levels. Various item-based service levels have been addressed in the literature. They can be defined in several ways, all of which are criticized by Chen (1996). Those of order-based levels are much more difficult to handle, and hence only a few studies have been reported (Hausman et al., 1998; Song, 1998). Moreover, in studies concerning service levels, some authors emphasize system evaluation, whereas others develop procedures for system optimization. These two approaches are clearly interrelated. This brief review does not distinguish between the two approaches, and focuses on item-based service levels.

A number of researchers have addressed the off-shelf ready rate and fill rate. Yano (1985) and Platt et al. (1997) deal with the re-order point, order quantity (i.e., (r, Q)) models with a fill-rate constraint. Rosling (2001a) represents the latest effort. He proposes optimal and near-optimal algorithms for two sets of problems. These algorithms are easy to implement. In the first set, he develops algorithms for optimizing several single-item inventory models with non-convex backlogging costs. However, he only provides a heuristic for finding the optimal (s, S) policies. In the second set of problems, Rosling considers service-level-constrained models, in which randomized policies may be optimal. Although one problem in his paper is similar to ours, he does not provide a procedure for optimizing it.

In an earlier paper, Rosling (2001b) identifies conditions under which non-convex backlogging costs can still result in quasi-convex loss functions. These costs are readily linked to several service-level constraints, including the off-shelf ready rate (i.e., T = 0). If the loss function is quasi-convex, we can then iteratively update the Lagrangian multiplier to find an optimal (s, S) policy that minimizes the cost and meets the service requirement. With each given value of the Lagrangian multiplier, we can apply the algorithm of Zheng and Federgruen (1991), or that of Feng and Xiao (2000), for the model optimization. However, this method is prone to degeneracy (Lau et al., 2002). “Degeneracy” signifies any situation in which the model cannot be solved and leads to nonsensical “optimal” solutions. In particular, there may exist a threshold level for the value of the Lagrangian multiplier, below which the optimal policy is meaningless and the achieved service level is zero, and above which an optimal policy is reasonable but the service level may be much higher than the target. Computationally, this will lead to an intensive search in the neighborhood of the threshold value.

Recently, the notion of advance demand information has received considerable attention. Advance Demand Information (ADI) is obtained when customers place orders in advance of their future requirements. Similar to the time-window setting, ADI means that firms can fulfill customer orders at a later time. However, the major difference between the two settings is that customers with ADI will not accept an earlier delivery; i.e., if a demand is of type T—period advance and arrives in period t, then the firm can ship it only in period t + T, provided that there is sufficient stock. Several incentives to elicit advance demand information have been studied in the literature, including price discounts and/or priority service to customers that book early (Chen, 2001). A similar notion of differentiating customers on the basis of demand leadtimes is explored in Wang et al. (2002).

Hariharan and Zipkin (1995) explore the continuous review model with ADI, where the base-stock policy is assumed and supply leadtimes may be stochastic. The demand that arrives at time t is expected to be met after T periods. The basic finding is that demand leadtimes are the opposite of supply leadtimes. Gallego and Özer (2001) address the periodic review model with setup costs, and prove the optimality of a modified (s, S) policy. In their model, demand that arrives in a period consists of T + 1 types: for type 0, “off-shelf” fulfillment is expected, and for each of the other types, t = 1, . . . , T, a fulfillment of i periods is expected. Demand that cannot be met on time is backlogged, and a shortage cost is incurred. If only type T demand exists, then their model is compatible with ours, with the difference that the fulfillment window in our model is more flexible. Furthermore, their model assumes a backlogging cost, while ours considers a service-level constraint.

3. Model formulation and preliminaries

3.1. Notation and assumptions

We consider a single-item, periodic-review system which is controlled by an (s, S) replenishment policy. Inventory can be replenished at the beginning of each period. The sequence of events in any period is as usual: (i) inventory review; (ii) placement of new orders; (iii) receipt of replenishing deliveries; (iv) arrival of demand; (v) fulfillment of demand; and (vi) evaluation of costs incurred in the period. All events are assumed to occur at the beginning of any period.

Demands for the item are i.i.d. across time periods, and the demand process is stationary. Partial fulfillment of a
demand is allowed. It is assumed that all demands unsatisfied are fully backlogged, and that backlogged demands are filled on a first-come first-served basis. For the purpose of computation, we consider discrete demand distributions. We use the terms increasing and decreasing in their weak senses, non-decreasing and non-increasing, respectively.

3.1.1. Notation

\[ K = \text{setup cost, } K \geq 0; \]
\[ IP(t) = \text{inventory position at the beginning of period } t, \text{ after possible ordering}; \]
\[ D_t = \text{demand in period } t, \text{ with mean } \lambda; \]
\[ \Phi(d) (\text{or } \phi(d)) = \text{probability distribution (or mass) function of } D_t, \phi(d) = 0, \forall d < 0; \]
\[ \Phi^n(d) (\text{or } \phi^n(d)) = n\text{-fold convolution of } \Phi(d) (\text{or } \phi(d)); \]
\[ D[p, t) = \sum_{i=p}^{t-1} D_i, \text{ for } t > p; \]
\[ D[p, t] = \sum_{i=p}^{t} D_i, \text{ for } t > p; \]
\[ L = \text{constant replenishment leadtime}; \]
\[ T = \text{length of time window}; \]
\[ \hat{L} = L - T \text{ if } T \leq L; \]
\[ \hat{L}' = T - L - 1 \text{ if } T > L; \]
\[ s, S = \text{re-order point and order-up-to level}; \]
\[ s^*, S^* = \text{optimal re-order point and order-up-to level}; \]
\[ Q = S - s, \text{ order quantity}; \]
\[ \beta(s, S) = \text{time window } T \text{ ready rate under policy } (s, S), T \geq 0; \]
\[ \beta_0 = \text{target service level}; \]
\[ H(y) = \text{expected holding cost in period } t + L, \]
\[ \text{where } m(0) = (1 - \phi(0))^{-1}, \]
\[ m(j) = \sum_{l=0}^{j} \phi(l)m(j - l), \; j = 1, 2, \ldots, \]
\[ M(0) = 0, \; M(j) = \sum_{l=0}^{j-1} m(l), \; j = 1, 2, \ldots, \]

are the well-known renewal functions.
The following lemma related to \( m(j) \) is needed later, and is proved in Appendix, A.

**Lemma 1.** The renewal density \( m(j) \leq m(0), \forall j \geq 0. \)

3.3. Expressions of time-window ready rates

As will be shown shortly, the problem structures are significantly different in the following two cases: (i) \( T \leq L \) (i.e., the time window is not longer than the replenishment leadtime); and (ii) \( T > L \). Thus, we derive the expression of ready rate for each case separately.

**Case (i) \( T \leq L \):** When \( T \leq L \), the analysis is simple. The ready rate considered here is the percentage of periods when there is no stockout. Whether the demand arriving in period \( t \) can be fulfilled by period \( t + T \) depends on the inventory position of period \( t + T - L \): \( IP(t + T - L) = y \). Specifically, demand in period \( t \) can be filled completely if and only if:

\[ y \geq D[t + T - L, t]. \]

We thus have the dis-service level with respect to \( y \):

\[ B(y) = 1 - \beta^{L-T+1}(y), \text{ if } y \geq 0, \text{ or } 1, \text{ if } y < 0. \] (2)

Letting \( T = 0 \), we obtain an expression for the "off-shelf" ready rate with lead-time \( L \). For \( T > 0 \), the above probability gives the ready rate with an "effective leadtime" of \( \hat{L} \leq L \). Hence, under an \((s, S)\) policy, the long-run average time-window ready rate can thus be written as:

\[ \beta(s, S) = \frac{\sum_{j=0}^{S-1} m(j)[1 - B(S - j)]}{M(S - s)}. \] (3)

In deriving the ready rate, we implicitly assumed that even if the demand in period \( t \) is zero, a negative inventory level in period \( t + T \) would still be counted as a stockout occurrence. Two alternative definitions are discussed in Appendix B. Although their dis-service level expressions look more complex than Equation (2), the differences do not entail any substantial revision on the subsequent optimization procedures.

**Case (ii) \( T > L \):** In case (ii) we further consider two subcases: A: \( s \geq -1 \) and B: \( s < -1 \). When demand \( D_t \) arrives at the beginning of period \( t \), the ordering decision for the period has been made, and so the inventory position
\( IP(t) = y > s \). Clearly, if \( s \geq -1 \), then \( IP(t) \geq 0 \), and no demand will wait longer than \( L + 1 \) periods to be fulfilled. Hence, all demands arriving in period \( t \) or earlier will be satisfied by period \( t + T \geq t + L + 1 \), resulting in a 100% ready rate, i.e., \( B(y) = 0 \) under case (ii)A.

Now turn to case (ii)B: \( s < -1 \). Four situations require different treatments:

a. if \( 0 \leq y - D_t \), then \( B(y) = 0 \);
b. if \( s < y - D_t < 0 \) and \( y - D_t - D[t + 1, t + T - L] \leq s \), then \( B(y) = 0 \);
c. if \( s < y - D_t < 0 \) and \( y - D_t - D[t + 1, t + T - L] > s \), then \( B(y) = 1 \); and

d. if \( y - D_t \leq s \), then \( B(y) = 0 \).

Under case (ii)Ba: \( s < 0 \leq y - D_t \), so that \( B(y) = 0 \) is obvious. In case (ii)Bb. \( D_t \) can still be fulfilled completely by period \( t + T \) because at least one order is placed during \([t + 1, t + T - L]\), which will bring the inventory position back up to \( S \geq 0 \) by period \( t + T \). Thus, \( B(y) = 0 \) in case (ii)Bb. However, in case (ii)Bc no order is placed before period \( t + T - L \), so \( D_t \) may not be completely satisfied by period \( t + T \). In case (ii)Bd, an order will be made in the next period, and all demand will be filled before \( t + L + 1 \), so \( B(y) = 0 \).

Analyzing case (ii)Bc requires quantifying the probability:

\[
\text{Prob} \{ s < y - D_t - D[t + 1, t + T - L] \leq s < y - D_t < 0 \} = \text{Prob} \{ D[t + 1, t + T - L] \leq y - D_t < 0 \}.
\]

Given \( y > s \), we have the dis-service level:

\[
B(s, y) = \sum_{z = \max(y + 1, 1)}^{y - s - 1} \Phi^L(y - s - 1 - z) \phi(z),
\]

where \( \Phi^0() = 1 \), and \( s \) is introduced in the notation to make the dependence on re-order level explicit.

Now, with the dis-service rate \( B(s, y) \), \( s < y \leq S \) being determined, the long-run average ready rate \( \beta(s, S) \) can be written as identical to Equation (3).

4. Optimization

For convenience, we denote \( (s, S) \) as \( (Q, S) \), where \( Q = S - s \). When no confusion could arise, we use \( C(s, S) \) (or \( \beta(s, S) \)) and \( C(Q, S) \) (or \( \beta(Q, S) \)) interchangeably. We also define the term “feasible \( S \)” as any order-up-to level such that \( \beta(S - 1, S) \geq \beta_0 \).

4.1. Monotonicity of cost \( C(s, S) \)

Lemma 2 below is basic to what follows.

Lemma 2. The long-run average cost \( C(Q, S) \) is an increasing function of \( S \) with \( Q \) being fixed, and a decreasing function of \( Q \) with \( S \) being fixed.

Proof. That \( \Delta C_S = C(Q, S + 1) - C(Q, S) > 0 \) is easy to verify, because \( H(y) \) is increasing. Now consider \( \Delta C_Q = C(Q + 1, S) - C(Q, S) \):

\[
\Delta C_Q = \frac{K + \sum_{j=0}^{Q} m(j) H(S - j)}{M(Q + 1)} - \frac{K + \sum_{j=0}^{Q-1} m(j) H(S - j)}{M(Q)}.
\]

\[
= \frac{M(Q) m(Q) H(S - Q) - m(Q) [K + \sum_{j=0}^{Q-1} m(j) H(S - j)]}{M(Q) M(Q + 1)}.
\]

\[
\leq 0.
\]

The last inequality holds because \( H(S - j) \geq H(S - Q) \) for all \( j = 0, \ldots, Q - 1 \).

From the monotonicity we know that the smaller \( S \) or the larger \( Q \) becomes, the lower the long-run average cost will be. So for any given \( S \) in the constrained optimization problem, we just need to find the largest \( Q \) such that \( (Q, S) \) satisfies the service-level constraint.

Lemma 3. For any feasible \( S \), the optimal order quantity

\[
Q(S) = \max(Q : \beta(Q, S) \geq \beta_0).
\]

Lemma 3 says that if a feasible order-up-to level is given, it is easy to locate \( Q(S) \). However, as will be shown, \( \beta(Q, S) \) may not exhibit monotonicity in \( Q \) even when \( S \) is fixed. Therefore, to develop algorithms using these monotonicity properties, the bounds on \( S \) and \( Q \) need to be established. Because these bounds are case dependent, we separately discuss the two cases below.

4.2. Optimization algorithm: Case (i)

As \( \Phi^{L+1}(y) \) is increasing in \( y \), the time-window ready rate \( \beta \) possesses the following properties, the proof of which is again straightforward.

Lemma 4. \( \beta(Q, S) \) is increasing in \( S \) with \( Q \) being fixed, and decreasing in \( Q \) with \( S \) being fixed.

The next lemma establishes a lower bound for the optimal \( S \).

Lemma 5. An optimal order-up-to level satisfies:

\[
S^* \geq S \quad \text{where:}
\]

\[
S = \min(S \geq 0 : \Phi^{L+1}(S) \geq \beta_0).
\]

Proof. Suppose \( S < S \) then no matter what \( Q \) is, the service level:

\[
\beta(Q, S) \leq \beta(S - 1, S) = \Phi^{L+1}(S) < \beta_0.
\]

The constraint can never be satisfied, so \( S^* \geq S \).
The optimal cost for any given $S$, say $C(S)$, is bounded by $\bar{C}(S)$, where:

$$C(S) = \frac{K + \sum_{j=0}^{S-1} m(j)H(S-j)}{S + 1} \frac{\beta_0}{m(0)},$$

which is increasing in $S$ when $S \geq S_0$, and $S_0$ is defined as:

$$S_0 = \min \left\{ S : \sum_{j=0}^{S} m(j)(H(S+1-j) - H(S-j)) \right\}.$$  \hspace{1cm} (12)

**Proof.**

Let

$$f(S) = \frac{K + \sum_{j=0}^{S-1} m(j)H(S-j)}{S + 1},$$

and define

$$\Delta f(S) = f(S+1) - f(S),$$

$$= \frac{\sum_{j=0}^{S} m(j)(H(S+1-j) - H(S-j)) - H(S-j) - K}{(S+2)(S+1)}.$$  \hspace{1cm} (13)

Because $H(y)$ is increasing and convex, and $H(0) = 0$, it follows that:

$$(S + 1)[H(S+1-j) - H(S-j)] - H(S-j) - (j + 1)[H(S+1-j) - H(S-j)] + (j - 1)[H(S-j) - H(S-j)],$$

which goes to infinity as $S$ goes to infinity. Thus, we can always find such an $S_0$ that $\Delta f(S) \geq 0, \forall S > S_0$. \hspace{1cm} ■

With a lower bound on $C(S)$, we can establish a terminating condition as follows: with every order-up-to level $\hat{S} \geq S$, we find the optimal quantity $Q(\hat{S})$, then calculate $C(\hat{S})$ and update the minimum cost $C^*$ if $C(\hat{S})$ is smaller than the current minimum cost obtained. Once $\hat{S} > S_0$ and $C(\hat{S}) > C^*$, we know that $C(S) \geq \bar{C}(S) \geq C(\hat{S}) > C^*, \forall S \in [\hat{S}, \infty)$, and there is no need to evaluate any $S \geq \hat{S}$. Hence, if every $S (\leq \hat{S})$ has been evaluated, directly or indirectly, then the search can terminate after $\hat{S}$.

Now the algorithm for the case (i) is ready.

**Algorithm (i)**

**Step 0.** Initialization: let $S = \hat{S}$ and $C^*$ be an arbitrary large number.
Step 1. Search for the largest $Q(S) \in [Q(S), \hat{Q}(S)]$ such that
\[ \beta(Q(S), S) \geq \beta_0. \]
Denote it as $\overline{Q}(S)$.

Step 2. Calculate $C(S) = C(Q(S), S)$. If $C(S) < C^*$ then
$C^* = C(S), S^* = S$ and $Q^* = Q(S)$.

Step 3. If $S > S_0$ and $\overline{C}(S) > C^*$, then end. Otherwise, let
$S = S + 1$ and go to Step 1.

In Step 0, $S$ is defined in Lemma 5. The initial cost is set to
be any sufficiently large value. An initial order-up-to level $S$ is also set to equal $S$. In Step 1, the optimal $Q(S)$ is found. This, together with $Q(S)$ and $\hat{Q}(S)$, is based on Lemma 6. Step 2 upgrades the optimal policy $(S^*, Q^*)$ and minimum cost $C(S)$ if the policy $(S, Q(S))$ identified in Step 1 achieves a smaller than previously obtained minimum cost. The algorithm terminates in Step 3 if $\overline{C}(S) > C^*$ (Lemma 7). Otherwise, the algorithm will take a bigger $S$ and returns to Step 1. Note that $S_0$ and $S$ can be obtained “off-line,” i.e., they can be calculated before the above algorithm is actually executed.

We also remark on the complexity of the algorithm. It is well known from the literature that the computational effort depends mainly on the number of $S$ being evaluated. This is because evaluating policy $(s, S)$ requires marginal effort after policy $(s, S)$ has been evaluated (Zheng and Fedewerken, 1991). Therefore, the complexity of the algorithm is proportional to the number $(S - S)$, where $S$ is defined as the largest $S$ being evaluated in the above algorithm.

4.3. Optimization algorithm: Case (ii)

The logic is the same as in case (i). Now, $S = 0$ because $(Q, S) = (1, 0)$ is always a feasible policy (i.e., $\beta(1, 0) = 1$). In case (i), $\beta(Q, S)$ is monotonic in $Q$ for any fixed $S$. However, this property does not hold anymore in case (ii). Lemma 3 offers little help in finding $Q(S)$ unless we have an upper bound for $Q(S)$. Such an upper bound, $\hat{Q}(S)$, is established in the ensuing lemma, and is used as the starting point of a one-dimensional descending search.

Lemma 8. For any given: $S \geq 0$:
\[ S + 2 \leq Q(S) \leq \hat{Q}(S), \]
where
\[ \hat{Q}(S) = \max\{Q : M(Q) \leq \frac{1 - \phi(0)}{\beta_0 - \phi(0)} [L' + M(S + 1)]\}, \]
and $\hat{Q}(S)$ is increasing in $S$.

The proof is lengthy and hence provided in Appendix C. Now we construct a terminating condition as in Lemma 7 in case (i). The proof is provided in Appendix D.

Lemma 9. Fix $S$, the optimal cost $C(Q, S)$ is bounded from below by $\underline{C}(S)$, where:
\[ \underline{C}(S) = \frac{K + \sum_{j=0}^{S-1} m(j)H(S-j)\beta_0 - \phi(0)}{L' + (S+1)m(0) (1 - \phi(0)).} \]

Furthermore, $\underline{C}(S)$ is increasing when $S \geq S_0$, where:
\[ S_0 = \min \left\{ S : L' \sum_{j=0}^{S-1} m(j)[H(S+1-j) - H(S-j)] - m(0)K > 0 \right\}. \]

As all the bounds are ready, so is the algorithm for case (ii).

Algorithm (ii)

Step 0. Initialization: let $S = 0$, and $C^*$ be an arbitrary large number.

Step 1. Let $Q = \hat{Q}(S)$, if $\beta(Q, S) < \beta_0$, $Q = Q - 1$ and re-evaluate $\beta(Q, S)$ until it reaches $\beta_0$.

Step 2. Calculate the cost $C(S) = C(Q, S)$, if $C(S) < C^*$ then $C^* = C(S), S^* = S$ and $Q^* = Q$.

Step 3. If $S > S_0$ and $\overline{C}(S) > C^*$, then end. Otherwise let $S = S + 1$ and go to Step 1.

In words, Step 1 works as follows. For any given $S$, it searches downward from the upper bound $\hat{Q}(S)$ to the first $Q$ such that the ready-rate constraint is met, because $C(Q, S)$ is decreasing in $Q$ with $S$ being fixed (see Lemma 2). Step 2 first computes the cost and then tests to see if the policy parameters and the minimum cost can be upgraded. Step 3 first tests if the search can be stopped (by Lemma 9). Otherwise, it continues with a bigger $S$ and returns to Step 1.

Suppose that the largest order-up-to level $S$ being evaluated in Step 3 is $\hat{S}$. Unlike case (i), an evaluation of policy $(s, S)$ and that of policy $(s, S)$ may be considerably different in case (ii). Hence, the computational effort of Algorithm (ii) is proportional to the number of $(s, S)$ policies being evaluated, which is bounded by $\hat{S} \times (\hat{Q}(\hat{S}) - \hat{S})$. Consequently, the algorithm’s complexity is roughly proportional to $\hat{S} \times (\hat{Q}(\hat{S}) - \hat{S})$.

5. Numerical examples

5.1. Cost-service trade-off

A time-window ready rate constraint consists of two parameters: the window span $T$ and the target rate $\beta_0$. A flexible fulfillment window allows a firm to reduce its inventories and better consolidate replenishment orders. In this section we use numerical examples to illustrate the trade-offs highlighted earlier.

In all examples, we assume the linear holding cost $H(y) = 0, \forall y \leq 0$, and $H(y+1) - H(y) = h\phi^{y+1}(y)$ for all $y > 0$. The single-period demand distribution is Poisson. We run a total of 7920 examples with the following parameter settings:

- leadtime $L = 1, 3, 5$
- time window $T = 0, 1, \ldots, 10$
unit holding cost \( h = 1, 3, 5; \)
setup cost \( K = 24, 44, 64, 84; \)
demand rate \( \lambda = 5, 10; \)
target ready rate \( \beta_0 = 0.50, 0.55, \ldots, 0.95. \)

In the following we evaluate the impacts of the following parameters: target ready rate \( \beta_0 \), time-window length \( T \), unit holding cost \( h \) and ordering cost \( K \). We also report on additional examples with \( \beta_0 = 0.96, 0.97, \ldots, 0.999 \).

5.1.1. Ready rate cost impact
First, we examine the trade-off between the ready rate and cost under different (fulfillment) window settings. We have the following combination of parameter values which we refer to as the “base case”: \( h = 1 \), \( \lambda = 5 \), \( K = 24 \) and \( L = 5 \). Figure 1 plots the result for the instances in which \( T \) increases from zero to 10 with an increment of two, in conjunction with different values of \( \beta_0 \). (To explore the non-linearity, we have included \( \beta_0 = 0.96, 0.97, \ldots, 0.999 \).) The non-linearity is striking when \( T \leq 4 \) but much milder as we increase \( \beta_0 \). (Due to the discrete nature of inventory policies, the curves for the cases \( T > L \) are not well shaped when \( \beta_0 \) is close to one.)

5.1.2. Time window cost impact
Second, the cost impact of the window length is evaluated. As can be seen from Fig. 2 whose parameter setting is the same as in the previous figure, the window length and the cost exhibit a much stronger trade-off relationship under all required levels of ready rate. This might be explained by the “diminishing marginal effect”; the more of one thing, there is the less its effect. Here it means that further lengthening the window span will result in a smaller cost saving from each additional period.

This pattern is enhanced by more examples (see Figs. 3 and 4). In the two figures we compare the costs between different time-window settings. Letting \( C^*(T) \) be the optimal cost with window \( T \), we define:
\[
\Delta_{T_1, T_2} = \frac{C^*(T_1) - C^*(T_2)}{C^*(T_1)} \times 100\%, \quad T_2 > T_1,
\]
as the marginal saving of extending the time window from \( T_1 \) to \( T_2 \). Similarly \( \Pi_{\beta_1, \beta_2} \) is defined as the marginal cost saving when the target service rate decreases from \( \beta_2 \) to \( \beta_1 \). They are plotted as follows: holding all other parameters the same, we compare the optimal costs between the window lengths \( T \) and \( T + 1 \), \( T = 0, 1, \ldots, 9 \). The difference in cost is captured by \( \Delta_{T, T+1} \). Thus, there are nine values of \( \Delta_{T, T+1} \) for each setting (which corresponds to a given set of \( (L, h, h, K, \lambda, \beta_0) \) values). We next change the parameter setting and finally take the averages of \( \Delta_{T, T+1} \) over all the combinations of parameter values. The result is shown in Fig. 3, where each point in the curve represents the average over 720 pairs of examples. The frequency statistics are summarized in Fig. 4 which plots three histograms: \( \Delta_{0,1}, \Delta_{2,3} \) and \( \Delta_{6,7} \). Again, each histogram plot represents 720 pairs of examples. Evidently, although the marginal cost saving of extending the time window by one period is decreasing in the window length \( T \), it is generally significant. For example, extending from the off-shelf to one-period window can save anywhere between 15 to 65% (see \( \Delta_{0,1} \) in Fig. 4),
with an average of 34% (see Fig. 3); in further extending six to seven periods, the saving ranges from 10 to 17% (see $\Delta_{6,7}$ in Fig. 4), with an average of 13.5% (see Fig. 3).

The “marginal” ready rate cost saving is reported at the bottom of Fig. 4. The comparison is done in exactly the same way as for $T$. Relatively speaking, the ready rate cost impact is not as striking as that of window span.

5.1.3. Cost impact of other parameters
Because the window span has a strong impact on cost, we further examine how other parameters affect this trade-off: leadtime $L$, unit holding cost $h$, setup cost $K$ and required service level $\beta_0$ (see Fig. 5). For each plot, the parameter concerned is fixed at one level and all other parameters vary their values to form different instances. The average

![Fig. 2. Time window cost impact.](image)

![Fig. 3. Average marginal time window cost saving.](image)
is plotted. Three series of histograms are reported in the figure: $\Delta_{0,1}$, $\Delta_{2,3}$ and $\Delta_{6,7}$. Taking the example of $K = 44$, the average is taken over 190 examples with different values of $L$, $h$, $\lambda$ and $\beta_0$, whereas for the example of $\beta_0 = 0.7$, the average is taken over 72 examples combining all values of $L$, $h$, $K$ and $\lambda$.

In addition to the previously observed patterns, Fig. 5 suggests that $\Delta_{i,i+1}$ is increasing in $h$ and decreasing in setup cost $K$. The former is clearly intuitive, whereas the latter might be explained as follows: although the time-window setting allows flexible fulfillment and hence reduces the average inventory level, the length of the replenishment cycle cannot be reduced too much, since the latter is determined by the difference between $s$ and $S$ and the demand rate.

There is no clear pattern regarding the relation between $\Delta_{i,i+1}$ and $L$ or $\beta_0$.

5.2. Performance of the algorithms

As analyzed earlier, the complexity of Algorithm (i) depends mainly on the number of $S$ being searched, and that of Algorithm (ii) is related to the number $\bar{S} \times (Q(\bar{S}) - \bar{S})$. As an illustration, Table 1 reports the result and computational performances of the algorithms, which is based on the following parameter setting: $h = 1$, $\lambda = 5$, $K = 24$, $\beta_0$, $L$, and $T$ are specified in the table. In the table, $s$ and $S$ are the optimal policy values, whereas $C$ and $\beta$ are the corresponding cost and expected ready rate, respectively. Evidently, the

Fig. 4. Marginal cost impact of the time window and ready rate.

Fig. 5. Cost impact of other parameters.
bounds are all much larger than the actual optimal parameter values. For example, in case (i), under the constrained ready rate of 0.70, the upper bound on $S$ from Algorithm (i) is 117, and that on $Q$ is 169. When comparing with actual values of 35 and 31, we see that the gaps are quite large. Clearly, there is still much room for tightening these bounds.

6. Concluding remarks

In concluding, we discuss several possible extensions of our model. First, the basic algorithmic results continue to hold in many continuous review systems with fixed order leadtimes. This applies, for example, to models with compound Poisson demands. The holding cost and dis-service level can be derived using the procedure proposed by Zheng and Federgruen (1991). The resulting cost and service-level expressions are almost identical to their periodic-review counterparts. We refer to Zheng and Federgruen (1991) for a more detailed discussion and further references.

Second, we have assumed a constant supply replenishment rate $S$. A general situation might involve random leadtimes. The ready-rate expression can easily be derived provided stochastic leadtimes are i.i.d. and not crossing (Zipkin, 1999). Consider the case when the leadtime is a Poisson random variable, $L = 1, 2, \ldots$, with probability $e^{-k}(k)^{T}/L$ where $k$ is the expected leadtime. Under a given $(s, S)$ policy, conditioning on $L \geq T$ yields the same expression as Equation (2) for the dis-service level, $B(y | L)$; and conditioning on $L < T$ yields Equation (4) for $B(s, y | L)$. Then, we take an expectation of $B(y)$ over all $L$ values to yield the expected dis-service level. However, the optimization procedures developed in this paper cannot be applied directly to this extension, and hence further study is needed.

Third, the current model assumes only one fulfillment time window. However, many companies, especially online retailers, allow multiple response times. It would be interesting to extend our single-window model to a multiple-window model. To begin with, consider just two window options (say, $T_1$ and $T_2$, where $T_1 < T_2 \leq L$) that are offered to customers and the demand for each is $D_1$ or $D_2$ per period, drawn from a given distribution. Conditioning on the inventory position of period $t + T_1 - L$ and following the steps in deriving Equation (2), it is not difficult to derive the two ready rates. Suppose that the inventory position in period $IP(t + T_1 - L) = y$. The dis-service level for the demand to be fulfilled within the $T_1$ window is the same as in Equation (2), with $T$ being replaced by $T_1$. However, the dis-service level for the demand to be fulfilled within the $T_2$ window depends on the inventory position in period $t + T_2 - L$, whose distribution is determined jointly by $y$ and the total demand during periods $t + T_1 - L$ to $t + T_2 - L - 1$. We can then derive the dis-service level for the demand with the $T_2$ window by conditioning on the inventory position in period $t + T_2 - L - 1$. We will now have two constraints: one for $T_1$, and the other for $T_2$. Quantities such as $m(t)$ and $M(t)$ will be calculated based on $D_1 + D_2$. We consider this model and its optimization an interesting topic for further research.

Finally, another possible direction for future research is the item-based time-window fill rate, which measures the percentage of demanded quantity fulfilled within the specified time span $T$. We still need to differentiate between two cases $T \leq L$ and $T > L$. In the case of $T \leq L$, the expression is similar to that of the off-shelf fill rate. The unfilled part of the demand that arrives in period $t$ is:

$$b(y) = E[(D[t + T - L, t] - y)^+] - (D[t + T - L, t] - y)^+,$$

in expectation, where $(x)^+ = x$ if $x \geq 0$, and = 0 otherwise. As for the case where $T > L$, we have a part of the demand arriving in period $t$ unfilled if and only if case (ii)BC occurs, because all the other cases will lead to a 100% fulfillment. Then, by the derivation of Equation (4), we can write the expected unfilled quantity as:

$$b(s, y) = \sum_{z = \max(y, 1, 1)}^{y - s - 1} z^\Phi(y - s - 1 - z)\phi(z).$$

---

**Table 1. Performance of the exact algorithms**

<table>
<thead>
<tr>
<th>$\beta_0$</th>
<th>$(S, \hat{S})$</th>
<th>$(Q(S), \hat{Q}(\hat{S}))$</th>
<th>$s$</th>
<th>$S$</th>
<th>$C$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50</td>
<td>(15,117)</td>
<td>(103,238)</td>
<td>8</td>
<td>34</td>
<td>3.519</td>
<td>0.506</td>
</tr>
<tr>
<td>0.55</td>
<td>(15,117)</td>
<td>(103,216)</td>
<td>5</td>
<td>35</td>
<td>3.865</td>
<td>0.553</td>
</tr>
<tr>
<td>0.60</td>
<td>(16,117)</td>
<td>(102,198)</td>
<td>1</td>
<td>34</td>
<td>4.176</td>
<td>0.600</td>
</tr>
<tr>
<td>0.65</td>
<td>(16,117)</td>
<td>(102,182)</td>
<td>2</td>
<td>34</td>
<td>4.539</td>
<td>0.652</td>
</tr>
<tr>
<td>0.70</td>
<td>(17,117)</td>
<td>(101,169)</td>
<td>4</td>
<td>35</td>
<td>4.904</td>
<td>0.701</td>
</tr>
<tr>
<td>0.75</td>
<td>(18,117)</td>
<td>(100,158)</td>
<td>7</td>
<td>33</td>
<td>5.257</td>
<td>0.753</td>
</tr>
<tr>
<td>0.80</td>
<td>(18,118)</td>
<td>(101,149)</td>
<td>9</td>
<td>33</td>
<td>5.654</td>
<td>0.806</td>
</tr>
<tr>
<td>0.85</td>
<td>(19,119)</td>
<td>(101,141)</td>
<td>10</td>
<td>36</td>
<td>6.066</td>
<td>0.851</td>
</tr>
<tr>
<td>0.90</td>
<td>(20,121)</td>
<td>(102,135)</td>
<td>12</td>
<td>37</td>
<td>6.635</td>
<td>0.902</td>
</tr>
<tr>
<td>0.95</td>
<td>(22,124)</td>
<td>(103,131)</td>
<td>15</td>
<td>35</td>
<td>7.301</td>
<td>0.951</td>
</tr>
</tbody>
</table>
The corresponding fill rate can then be obtained by dividing $b(y)$ or $b(s, y)$ by $E(D_s)$. Although this model is similar in spirit to our ready-rate model, its optimization will call for a different approach. This is another interesting topic for future research.

Acknowledgements

The author list is randomly ordered, and each contributes equally to the paper. We appreciate the three referees’ comments, which have helped us improve the exposition of the paper. We are also grateful to Professor Xiuli Chao of North Carolina State University for his thorough reading of this paper. He not only pointed out several technical oversights in the earlier version, but also suggested ways to correct them.

References


Feng, Y. and Xiao, B. (2000) A new algorithm for computing optimal $(s, S)$ policies in a stochastic single item/location inventory system. IIE Transactions, 32(11), 1081.


Appendices

Appendix A: Proof of Lemma 1. By induction. Clearly when $j = 1$, $m(1) = m(0)\phi(1)/1 - \phi(0) \leq m(0)$ because $\phi(1) \leq 1 - \phi(0)$. Now suppose that $m(j) \leq m(0)$, $\forall j \in [0, k]$, then.

$$m(k + 1) = \frac{\sum_{j=0}^{k} m(j)\phi(k + 1 - j)}{1 - \phi(0)} \leq m(0) \cdot \frac{\sum_{j=0}^{k} \phi(k + 1 - j)}{1 - \phi(0)},$$

$$= m(0) \cdot \frac{\sum_{j=1}^{k+1} \phi(j)}{\sum_{j=1}^{\infty} \phi(j)},$$

$$\leq m(0).$$

Therefore, $m(j) \leq m(0)$ for all positive-integer $j$.

Appendix B: Alternative definitions of ready rate

There are different definitions of the ready rate according to the differing treatments of the zero demand cases. One alternative is to assume that periods with zero demand are completely fulfilled no matter what the inventory in period

$$p_{\text{IP}(T - L) \geq D[T - L, 0] | D_0 > 0},$$

$$= \frac{p_{\text{IP}(T - L) \geq D[T - L, 0]} - p_{\text{IP}(T - L) \geq D[T - L, 0] | D_0 = 0} p_{D_0 = 0}}{p_{D_0 > 0}},$$

$$= \frac{\Phi^{-L+1}(y) - \Phi^{-L}(y)\phi(0)}{1 - \phi(0)}.$$
$t + T$. For simplicity, set $t = 0$ in what follows. The probability of no stockout can be expressed as a function of $IP(T - L) = y$:

$$P(IP(T - L) \geq D[T - L, 0] | D_0 > 0) P(D_0 > 0) + 1 \times P(D_0 = 0)$$

$$= P(IP(T - L) \geq D[T - L, 0] | P(IP(T - L)$$

$$\geq D[T - L, 0] | D_0 = 0) P(D_0 = 0) + 1 \times P(D_0 = 0),$$

$$= P(IP(T - L) \geq D[T - L, 0]) - P(IP(T - L)$$

$$\geq D[T - L, 0]) P(D_0 = 0) + 1 \times P(D_0 = 0),$$

$$= \Phi^{T-L+1}(y) - \Phi^{T-L}(y) \phi(0) + \phi(0).$$

Another definition can be viewed as the percentage of periods with positive demand that are completely fulfilled within the time window. The respective probability is then:

$$\Phi^{T_L}(y) - \Phi^{T_L}(y) \phi(0) = \sum_{j=1}^{y} \Phi^{L-T}(y - j) \phi(j),$$

is increasing in $y$. Moreover, these definitions satisfy the properties developed in this paper, and similar properties can be developed in the same manner.

Appendix C: Proof of Lemma 8. Denote by $Q$ the set $\{ Q : \beta(Q, S) \geq \beta_0 \}$. Clearly, only the elements in set $Q$ can satisfy the service constraint. Therefore, $Q(S)$ is the one that minimizes $C(Q, S)$. From Lemma 2, we know $C(Q, S)$ is decreasing in $Q$ with $S$ being fixed. Therefore, $Q(S)$ is the maximal element in set $Q$.

To prove the upper bound of $Q(S)$, we need to simplify $\beta(s, S)$ first.

$$S = \sum_{j=0}^{S-1} m(j) \beta(S - j)$$

$$= \sum_{j=0}^{S-1} \sum_{z=(S-j)+1}^{S-1} \Phi^{L}(S - s - 1 - j - z) \phi(z),$$

$$= \sum_{j=0}^{S-1} \sum_{z=0}^{S-j} \Phi^{L}(S - s - 1 - j - z) \phi(z)$$

$$- \sum_{j=0}^{S-1} m(j) \sum_{z=0}^{S-j} \Phi^{L}(S - s - 1 - j - z) \phi(z),$$

$$= \sum_{j=0}^{S} m(j) \Phi^{L}(S - s - 1 - j) - \Phi^{L}(S - s - 1)$$

$$- \sum_{j=0}^{S} m(j) \Phi^{L}(S - s - 1 - j) \phi(z)$$

$$- \sum_{j=S+1}^{S} m(j) \Phi^{L}(S - s - 1 - j) \phi(0),$$

$$= \sum_{j=0}^{S} m(j) \Phi^{L}(S - s - 1 - j) - \Phi^{L}(S - s - 1)$$

$$- \sum_{j=S+1}^{S} m(j) \Phi^{L}(S - s - 1 - j) \phi(0).$$

The service-rate constraint can be re-written as:

$$\beta(s, S) = 1 - \frac{\sum_{j=0}^{S-1} m(j) \Phi^{L}(S - s - 1 - j) (1 - \phi(0))}{\sum_{j=0}^{S-1} m(j) \Phi^{L}(S - s - 1 - j)} \geq \beta_0.$$  \hspace{1cm} (16)

When the inequality in Equation (17) does not hold, the required service rate $\beta_0$ cannot be reached. When $S$ is given, the right-hand side is a constant, which is larger than $M(S + 1)$. So when $Q = S + 1$, the inequality in Equation (17) always holds. The left-hand side is increasing in $Q$ and goes to infinity as $Q$ goes to infinity. Thus, a $Q(S)$ must exist.

Appendix D: Proof of Lemma 9. Notice that $H(s + 1, \ldots, H(0) = 0$. By the inequality in equation (17):

$$C(Q, S) \leq \frac{K + \sum_{j=0}^{S-1} m(j) H(S - j)}{M(Q(S))} \geq \frac{K + \sum_{j=0}^{S-1} m(j) H(S - j)}{L' + \frac{S}{(S+1)} m(0) - 1 - \phi(0)}.$$
the following derivation. Letting:

\[ f(S) = \frac{K + \sum_{j=0}^{S-1} m(j)H(S - j)}{L' + (S + 1)m(0)}, \]

we only need to prove that \( f(S) \) is increasing when \( S \geq S_0 \).

\[
\Delta f(S) = f(S + 1) - f(S) = \frac{K + \sum_{j=0}^{S} m(j)H(S + 1 - j) - \sum_{j=0}^{S} m(j)H(S - j)}{L' + (S + 1)m(0)}.
\]

\[
\Delta f(S) = \frac{L' \sum_{j=0}^{S} m(j)(H(S + 1 - j) - H(S - j)) - m(0)K + m(0) \sum_{j=0}^{S} m(j)(S + 1)\{H(S + 1 - j) - H(S - j)\} - H(S-j))}{L' + (S + 1)m(0)}.
\]

Let \( S_0 = \min\{S': \sum_{j=0}^{S} m(j)(H(S + 1 - j) - H(S - j)) - m(0)K > 0\} \). Because \( H(y) \) is an increasing and convex function in \( y \), and \( \lim_{y \to -\infty} H(y) = \infty \), the first term in the numerator of the above inequality will go to infinity as \( S \to \infty \). Thus, the existence of \( S_0 \) is ensured. When \( S \geq S_0 \), the lower bound \( \zeta(S) \) is increasing in \( S \) and goes to infinity as \( S \to \infty \).

Biographies

Tong Wang is currently a Ph.D. student in the Technology Management area, INSEAD, France. He received his M.Phil degree in Systems Engineering from The Chinese University of Hong Kong and his B.E. degree in Industrial Engineering from Shanghai Jiao Tong University. His main research interests include sales effort and inventory control, and inventory models.

Frank Y. Chen has been an Associate Professor since 2001 in the Department of System Engineering and Engineering Management, the Chinese University of Hong Kong. He received a B.E. degree in Mechanical Engineering and a M.A. degree in Economics from Tsinghua University (Beijing) and the University of Waterloo, respectively. After finishing his Ph.D. in Operations Management from the University of Toronto, Dr. Chen took a post-doctoral position at Northwestern University. During 1997–2001, he taught in the School of Business, National University of Singapore. His current research focuses on joint decision-making in sales effort and inventory control, and inventory models.

Youyi Feng received both his B.S. degree and M.S. degree in Mathematics from Nanjing University, and his Ph.D. degree in Operation Research from Columbia University. During 1995–2000, he was an assistant professor in the Department of Information Systems at National University of Singapore. Before he joined the Chinese University of Hong Kong as an Associate Professor in 2001, he had worked as a research manager in Enron North America. His research interests include revenue management, supply chain management, and optimal production control.

Contributed by the Applied Optimization Department