A new two-bin policy for inventory systems with differentiated demand classes

Ghosh, Sugoutam; Piplani, Rajesh; Viswanathan, S.

2014

https://hdl.handle.net/10356/101597

https://doi.org/10.1111/poms.12299

A New Two-Bin Policy for Inventory Systems with Differentiated Demand Classes

Sugoutam Ghosh

Nanyang Business School, Nanyang Technological University, 50, Nanyang Avenue, Singapore, 639798, Singapore
sugoutam@gmail.com

Rajesh Piplani

School of Mechanical & Aerospace Engineering, Nanyang Technological University, 50, Nanyang Avenue, Singapore, 639798, Singapore
mrpiplani@ntu.edu.sg

S. Viswanathan

Nanyang Business School, Nanyang Technological University, 50, Nanyang Avenue, Singapore, 639798, Singapore
asviswa@ntu.edu.sg

We consider an inventory system under continuous review with two demand classes that are different in terms of service level required (or penalty cost incurred for backordering of demand). Prior literature has proposed the critical level rationing (CLR) policy under which the demand from the lower priority class is backordered once inventory falls below the critical level. While this reduces the penalty cost for the higher demand class, the fill rate achieved for the lower priority demand class gets compromised. In this study, we propose a new class of two-bin (2B) policy for the problem. The proposed 2B policy assigns separate bins of inventory for the two demand classes. The demand for each class is fulfilled from its assigned bin. However, when the bin intended for the higher demand class is empty, the demand from the higher class can still be fulfilled with the inventory from the other bin. The advantage of the 2B policy is that better fill rates are achieved, especially for the lower demand class. Computational results show that the proposed policy is able to provide a much higher service level for the lower priority class demand without increasing the total cost too much and without affecting the service level for the higher priority class. When a service level constrained optimization problem is considered, the 2B policy dominates the CLR policy when the service level difference for the two classes is not too high or the service levels required for both the classes are relatively lower.

1. Introduction

In today's competitive environment, manufacturing firms are increasingly focusing on aftermarket services, where service differentiation is seen as a key strategy to manage costs and deliver what is promised to the customer. Inventory policies that can tackle differentiated service requirements effectively and efficiently are therefore very important in the management of service parts logistics. Many organizations segregate the demand for the same product into multiple customer classes or demand classes according to the different priorities for order fulfillment. (Throughout the study, we use the terms *demand classes*, *customer classes*, and *order classes* interchangeably to mean the same thing).

Our interactions with the regional warehouse of a leading automobile company distributing spare parts have shown that they use three order classes for the purposes of service differentiation: normal, urgent, and vehicle off-road (emergency). Normal orders (or demands) are fulfilled immediately on arrival only if the inventory level is higher than the reserve inventory level; else they are backordered. The reserve inventory is kept aside for fulfilling urgent demand. If there is no inventory at all in the regional distribution warehouse, then urgent orders are also backordered and replenishment orders placed with the headquarters are expedited. Vehicle off-road (or emergency) orders are situations where there is demand for a very high-value spare part (whose failure has caused the vehicle to be off-road). Demand for these high-value spare parts occur with very low probability, and therefore are fulfilled from the central warehouse in the headquarters directly by emergency air-shipments. Similar observations were made in our interactions with an aircraft service company. In military materials management, when the same spare part is

requisitioned from different divisions, an inventory control policy that prioritizes the orders according to the end use need to be deployed so that the more critical demand is fulfilled first (Deshpande et al. 2003).

Inventory rationing has been shown to be an effective strategy when there are multiple demand classes requiring different service levels (or incurring different penalty costs). Critical level rationing (CLR), where the lower class demand is either backordered or not satisfied after the on-hand inventory reaches a pre-determined critical level (called rationing level) is the most widely studied and recommended policy in the available literature.

Although CLR provides an inventory policy with a lower cost (Deshpande et al. 2003), its major shortcoming is that by reserving inventory, it provides a high service level to the higher priority customer class at the expense of the lower priority class. It is conceivable that the inventory reserved for the higher priority demand class may not get fully utilized if there is only little demand for the higher class after rationing kicks in. The lower priority demand class, therefore, may not be fulfilled on time even when there is inventory available. Thus the CLR policy results in a low fill rate for the lower priority demand class. Numerical experiments in our study show that if no fill rate constraints are imposed, the service level or fill rate for the lower priority demand class suffers significantly in the CLR policy. For example, in some cases, the policy provides a fill rate of only 64% to the lower priority class. In this study, a new two-bin (2B) policy is proposed. This policy attempts to provide a higher service level to the lower priority demand class while maintaining the service level for the higher priority class.

In our proposed policy, two separate bins, BIN_1 and BIN_2, are kept for the two demand classes; with class 1 being the higher priority demand class. Demand from each class is satisfied from its respective bin. However, the demand from class 1 can be fulfilled with inventory from BIN_2 (if available), once BIN_1 becomes empty, on a first-come-first-serve (FCFS) basis. The lower priority demand class, however, cannot use BIN_1 when BIN_2 runs out. This pol-

icy ensures that the lower class demand is not backordered as long as there is inventory left in BIN_2. At the same time, it preserves the priority of the higher demand class by allowing it to use the lower class's stock once its own stock runs out. Thus this policy gives some protection to the lower demand class while still providing higher priority to class 1 demand. The policy is depicted in Figure 1.

Numerical experiments discussed later in the study show that, for the problem of minimizing the total (ordering + holding + penalty) cost with no restrictions on the demand fill rate, our proposed policy increases the service level for the lower class demand by as much as 25%, with only a slight increase in cost. When a service level constrained optimization problem is considered, where instead of imposing penalty costs, fill rate constraints are imposed on the two demand classes and the total cost (comprised of ordering and holding cost) is minimized, our proposed policy provides a lower cost than the CLR policy in many instances.

The remainder of this paper is organized as follows. We first provide a brief review of the related literature at the end of this section. In section 2, the proposed 2B policy for inventory rationing is explained. The model for this policy is formulated and the exact expression for the expected cost is developed. The algorithm to determine the optimal parameters is described next. Section 3 provides the results of computational experiments comparing the proposed 2B policy with the CLR policy. The final section provides concluding remarks and a summary.

1.1. Literature Review

Over the years many researchers have studied inventory rationing. Veinott (1965) was the first to propose inventory rationing but Topkis (1968) first analyzed, theoretically, the concept of inventory rationing and developed allocation policies for inventory among multiple classes. Evans (1968) and Kaplan (1969) independently derived the same results as Topkis (1968) for two customer classes. Evans (1968) assumed that unsatisfied demand is lost, whereas Kaplan (1969)

Figure 1   Proposed Two-Bin (2B) Policy
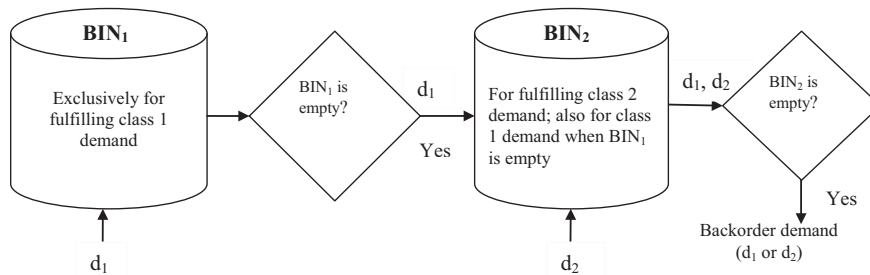
(d_1 is class 1 demand, and d_2 is class 2 demand)

assumed that it is backordered to the next cycle. These studies considered period review inventory systems with zero lead time and the focus was on developing the structure of the optimal inventory and rationing policy.

Nahmias and Demmy (1981) were the first to consider continuous review inventory systems under Poisson demand. In their study, they first considered a single period model, and then developed an approximate continuous review model with the assumption that there is never more than one order outstanding. Their focus was on developing and evaluating approximate expressions for fill rates for a given inventory policy and rationing level, rather than on optimizing the cost. Later, Moon and Kang (1998) extended the same idea to compound Poisson process, but using a simulation model. Some researchers such as Evans (1968), Dekker et al. (2002), and Melchiors et al. (2000), considered the lost sale case, where the unmet demand is not satisfied and assumed as lost, incurring a one-time penalty cost. Lost sale is more appropriate in a retail setting where rationing might not always be appropriate or practical. In industrial or B2B settings, such as spare parts distribution, it is more appropriate to assume that demand not satisfied immediately is backordered, rather than lost.

Recent work on inventory rationing in a continuous review environment with setup cost includes Deshpande et al. (2003) and Arslan et al. (2007). Deshpande et al. (2003) considered an inventory system with two customer classes whose demands follow Poisson processes. The inventory is replenished according to a (Q, r) policy where a replenishment order of quantity Q is placed whenever the inventory position drops to level r. The demand is fulfilled from inventory on a FCFS basis, as long as on-hand inventory is equal to or above the rationing level, K. Demand from the lower priority class (class 2) is backordered once on-hand inventory falls below the rationing level, but demand from higher priority class (class 1) is still satisfied as long as there is on-hand inventory. Demand from the higher priority class is backordered only after the inventory runs out. Backorders are cleared when the next replenishment arrives according to a *threshold clearing mechanism* that they proposed. The threshold clearing mechanism clears backorders considering inventory position rather than inventory level. This means that all the lower class demand that occurs after the placement of replenishment order, until the $(r + Q - K)$th demand arrival, is cleared upon receipt of the replenishment. The threshold clearing mechanism aids in the regeneration of the inventory positions and thus facilitates the analytical derivation of the probability distribution of inventory levels easily as compared to *priority clearing*, where

backorders are cleared based on the priority of the demand. Therefore, accurate expressions for the expected cost can be developed. For a (Q, r) policy, with a fixed rationing level, K, Deshpande et al. (2003) derived the expression for long-run expected cost. Using the algorithm of Federgruen and Zheng (1992), the optimal (Q, r) policy for a fixed K is determined; the optimal rationing level can then be determined by exhaustive search over all possible rationing levels. Deshpande et al. (2003) also showed that the simulated cost for the priority clearing mechanism using the optimal policy parameters obtained by assuming threshold clearing was not very different from the cost for the threshold clearing mechanism.

Arslan et al. (2007) considered the same clearing mechanism as Deshpande et al. (2003) but considered 'n' customer classes. They developed an equivalent serial-stage inventory system framework for solving the problem. In their problem formulation, the objective was to minimize the expected average inventory level subject to constraints on the pre-defined service levels for the different customer classes.

Another recent work on spare parts inventory management (in a multi-item, multi-location network) is by Alvarez et al. (2013). They propose a new strategy where dedicated stock is kept at the customers' site along with a common stock at a central location. The concept of dedicated stock might seem similar to the bin stock under our proposed 2B policy, but it is not. Under the 2B policy, the higher class demand can be fulfilled from BIN_2, if BIN_1 is empty. In the policy by Alvarez et al. (2013), the dedicated stock at one location cannot be used to fulfill demand at another location. All the customer classes can use the common stock only if its own dedicated stock is empty. So in a sense, the policy by Alvarez et al. (2013) is more akin to a modified critical level policy. Moreover, their study does not consider ordering cost, and therefore lot for lot ordering is pursued; this makes the computation of the holding cost very straight forward. Furthermore, their cost evaluations are only approximate.

All the above studies and the problem being considered in this study assume that it is an inventory system with fixed lead time for replenishment. It is also relevant to point out there is a body of literature considering single product manufacturing systems with multiple demand classes under a make-to-stock framework. Studies in this body of literature do not consider any fixed setup cost, and use a queuing theoretic framework to model the lead time for orders that are queued in the system. The focus of these studies is to analyze the structure of the optimal policy for fulfilling/rationing of the incoming orders. Studies belonging to this category include Ha (1997a,b,

2000), de Véricourt et al. (2000, 2002), and Gayon et al. (2009).

2. The Proposed Two-Bin Policy

The notation used in the study is presented in Table 1. We consider a single item, continuous review inventory system that follows a (Q, r) policy, wherein a replenishment order of quantity Q is placed when the inventory position (inventory on hand + inventory on order − backorders) of the system drops to r. The demand for the item comes from two different customer classes that vary in their penalty cost or fill rate requirement. The demand from each customer class, i, follows a Poisson process with intensity λ_i; the total demand from two classes therefore also follows a Poisson process with intensity $\lambda = \lambda_1 + \lambda_2$. Each replenishment order incurs a fixed ordering cost A. Holding cost is incurred at the rate of h per unit of on-hand inventory per unit time. The (fixed) lead time for replenishment is L (>0). The demand that is not satisfied immediately upon arrival is backordered and incurs two types of penalty costs: a delay cost at rate p_i per unit per unit time for the duration the demand is not fulfilled, and a one-time stock-out cost of π_i per unit. It is assumed that class 1 demand has higher priority than class 2 demand; accordingly the penalty cost for class 1 is higher than that of class 2, that is, $p_1 \geq p_2$ and $\pi_1 \geq \pi_2$.

Two separate bins are kept for the two customer classes, BIN_1 and BIN_2. As a (Q, r) policy is followed and the demand follows a Poisson process, the total inventory position immediately after an order is placed is $r + Q = S_1 + S_2$, where S_1 and S_2 are the base stocks allocated to BIN_1 and BIN_2, respectively. The quantity ordered for each bin (denoted by q_1 and q_2, respectively) is equal to the base stock allocated to that bin minus its inventory position at the time of ordering. Note that, $Q = q_1 + q_2$.

The 2B policy proposed in this study works as follows. Demand that arrives for each class is fulfilled from their respective bins, if possible. When BIN_2 is empty, demand for class 2 is backordered. When BIN_1 is empty, demand for class 1 is fulfilled from BIN_2, if sufficient inventory is available in BIN_2, if not, it is backordered (see Figure 1). We assume that the backorders are cleared using the threshold clearing mechanism, in order to develop the exact cost expression for the 2B policy. Under threshold clearing, backorders in the current order cycle are cleared when replenishments arrive, if the inventory position in the respective bins was positive at the time of the corresponding demand arrival. If not, they are cleared from the replenishment in the next order cycle. The threshold clearing mechanism ensures that the inventory positions are regenerated in every replenishment cycle and thus the holding and penalty costs can be calculated exactly. This is unlike the priority clearing mechanism, where the backorders for the higher demand class is always cleared first when replenishment arrives. Therefore, as in Deshpande et al. (2003), we use the threshold clearing mechanism to develop the optimal 2B policy and to evaluate the costs, even though priority clearing is easier to implement in practice. As discussed earlier, Deshpande et al. (2003) had shown that the expected cost obtained by the priority clearing mechanism is close to that of the threshold clearing mechanism in simulation experiments. A more detailed explanation of the implementation of the threshold clearing mechanism for the 2B policy is provided in the unabridged version of the study.

In practice in an ERP system, the 2B policy with priority clearing can be implemented by treating the two bins as the separate items (or same item in two different locations), with the additional condition that item 1 can be substituted by item 2. Furthermore, the reorder should be triggered only based on the sum of the inventory positions of the two items. On the basis of the conversation with a leading ERP system provider, we believe that the 2B system can be implemented by customizing the ERP system. If not, a special purpose inventory software package has to run on top of the ERP system to implement the policy.

2.1. Evaluation of The Policy Cost

The decision variables for the proposed policy are the order quantity Q and the base stock positions of BIN_1 and BIN_2, S_1 and S_2, respectively. Note that the reorder point, $r = (S_1 + S_2 - Q)$. Let $C(Q, S_1, S_2)$ denote the long-run expected cost of the policy for a given Q, S_1, and S_2. The expected cost comprises of three components—ordering cost, holding cost, and penalty cost. Since the order quantity is Q for every order, the expected ordering cost is $\frac{A\lambda}{Q}$. Therefore, the expected cost can be written as

$$C(Q, S_1, S_2) = \frac{A\lambda}{Q} + G(Q, S_1, S_2), \qquad (1)$$

Table 1 Notation

i	Index for demand class, $i = 1, 2$
λ_i	Demand rate for class i
λ	Total demand rate for the two classes. $\lambda = \lambda_1 + \lambda_2$
p_i	Time-weighted delay cost per unit for class i
π_i	Fixed stock-out cost per unit for class i
h	Holding cost per unit per unit time
A	Fixed ordering cost per order
L	Lead time for replenishment arrival
Q	Order quantity
r	Reorder point
S_i	Base stock for class i

where $G(Q, S_1, S_2)$ is the expected holding and penalty cost for a given policy.

Given that a (Q, r) policy is used, the total inventory position of the two bins follows a regenerative process. Let the lth order be placed at time τ_l and the $l + 1$th order at τ_{l+1}. Let t be any point in the lth replenishment cycle such that $\tau_l < t \leq \tau_{l+1}$ (see Figure 2). Knowing the probability distribution of inventory position at time t (or total demand in the interval $(\tau_l, t]$), and the demand during period $(t, t+L]$, the cost $C(Q, S_1, S_2)$ can be evaluated. The order placed at τ_l would be received at τ_l+L. Therefore the inventory represented by inventory position (S_1, S_2) at τ_l would be fully available at τ_l+L, and this inventory position minus the demand during $(\tau_l, t+L]$ would accurately represent the inventory level/backorders at time $t+L$. Also, due to the design of the threshold clearing mechanism, the backorders from the previous replenishment cycles are already accounted for in the inventory position S_i (even though they may be cleared only at τ_l+L) and can be ignored for computing the inventory levels and cost at time $(t+L)$. However, the sequence of demand arrivals during the period $(\tau_l, t+L]$ is important in computing the correct penalty costs. Since a (Q, r) policy is used, and the demand follows a Poisson distribution, the total demand for the two classes, $\bar{D}^t$ during the period $(\tau_l, t]$ follows a Uniform distribution with probability

$$u(\bar{D}^t = i) = \frac{1}{Q}, \ i = 0, 1, 2 \ldots, Q - 1. \quad (2)$$

The total demand for the two classes during the period $(t, t+L]$, $\bar{D}^L$ follows a Poisson distribution,

$$\Pr(\bar{D}^L = j) = \frac{e^{-\lambda L}(\lambda L)^j}{j!} j = 0, 1, \ldots \infty. \quad (3)$$

Let $\bar{D}^{t+L}$ be the total demand during the period $(\tau_l, t+L]$. Then the probability distribution of $\bar{D}^{t+L}$ can be computed as a convolution of the Uniform and the Poisson distribution as follows

$$\Pr(\bar{D}^{t+L} = j) = f_{t+L}(j)$$
$$= \sum_{i=0}^{\min(j,Q-1)} \left(\frac{1}{Q} \frac{e^{-\lambda L}(\lambda L)^{j-i}}{(j-i)!} \right) j = 0, 1, \ldots \infty. \quad (4)$$

Figure 2 Timeline of The Replenishment Cycles
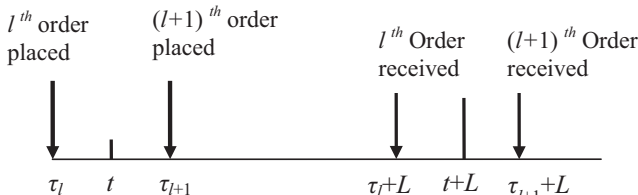
Since the demands follow Poisson process, the probability that a particular demand arrives from class i ($i = 1, 2$) is $\alpha_i = \frac{\lambda_i}{\lambda}$ and the number of ith class demand arrivals, D_i, out of a total demand, $\bar{D}^{t+L}$ follows a binomial distribution given by

$$B(D_i, \bar{D}^{t+L}, \alpha_i) = {}^{\bar{D}^{t+L}}C_{D_i} \alpha_i^{D_i} (1 - \alpha_i)^{\bar{D}^{t+L} - D_i}, \quad (5)$$

where, ${}^{\bar{D}^{t+L}}C_{D_i} = \frac{\bar{D}^{t+L}!}{D_i!(\bar{D}^{t+L} - D_i)!}$ is the number of ways of choosing D_i class i demands out of the total $\bar{D}^{t+L}$ demand arrivals. If k_1 and k_2 are the actual number of demand arrivals for class 1 and class 2, respectively, during the time period $(\tau_l, t+L]$, the probability distribution of k_1 and k_2 can be computed as a convolution of the Poisson distribution, given by Equation (4) and the binomial distribution, given by Equation (5). The joint probability distribution of demand for the two classes D_1 and D_2 during the time period $(\tau_l, t+L]$ is given by

$$g(k_1, k_2) = \Pr(D_1 = k_1, D_2 = k_2)$$
$$= f_{t+L}(k_1 + k_2)B(k_1, k_1 + k_2, \alpha_1). \quad (6)$$

As class 1 demand can also be satisfied from BIN_2, if BIN_1 is empty, the expected holding and penalty costs depend on the sequence of demand arrivals. Let $l_1 = S_1 - k_1$ and $l_2 = S_2 - k_2$. When both l_1 and l_2 are greater than zero, they represent the inventory levels in BIN_1 and BIN_2 at time $(t+L)$. In other cases, one or both demand classes can incur penalty costs. Table 2 provides a list of possibilities for the holding and penalty costs for different values of l_1 and l_2.

Cases (1) and (2) do not require class 1 demand to use BIN_2 because in both these cases, the inventory in BIN_1 is enough to satisfy all the demand from class 1. In Case (3a), BIN_1 runs out of inventory, but the remaining class 1 demand can be fulfilled from BIN_2. However, in cases (3b) and (4) the inventory available in BIN_2 is not enough to satisfy all the demand from the two classes. In these cases, the number of units of class 1 demand (and class 2 demand) that are backordered depends upon the sequence of demand arrivals. Let $G_j(Q, S_1, S_2)$ represent the expected holding and penalty cost for case j, $j = 1, 2, 3a, 3b, 4$.

Case 1: In this case, all demands for each class are satisfied from their respective bins and there are no backorders at time $(t + L)$. The units remaining after satisfying the demands incur holding cost; the expected penalty and holding cost per unit time for this case is

$$G_1(Q, S_1, S_2) = \sum_{l_1=1}^{\infty} \sum_{l_2=1}^{\infty} h(l_1 + l_2)g(S_1 - l_1, S_2 - l_2). \quad (7)$$

Note that for a given l_1, l_2, the realized demand for the two classes during the period $(\tau_l, t+L]$ is

Table 2. Effect on Holding and Penalty Cost for Different Cases of l_1 and l_2

Case	l_1	l_2		Description	Associated cost
1	>0	>0		Demands for both classes are fulfilled from their respective bins	Holding costs for both BIN_1 & BIN_2
2	>0	≤0		Demand for class 1 is fulfilled from its bin, BIN_2 runs out, and class 2 demands are backordered	Holding costs for BIN_1 & penalty cost for class 2 demand
3a	≤0	>0	$|l_1| < l_2$	All demands are satisfied, class 1 demand is partly fulfilled from BIN_2. BIN_1 has zero inventory	Holding cost for BIN_2
3b			$|l_1| \geq l_2$	Both BIN_1 and BIN_2 have zero inventory. Some of the demands are backordered	Penalty cost for class 1 and class 2 demand, depending on sequence of demand arrivals
4	≤0	≤0		Both BIN_1 and BIN_2 have zero inventory. Some of the demands are backordered	Penalty cost for class 1 and class 2 demand, depending on sequence of demand arrivals

$k_1 = (S_1 - l_1)$, and $k_2 = (S_2 - l_2)$, respectively. Correspondingly the joint probability distribution of the demand is $g(S_1 - l_1, S_2 - l_2)$.

Case 2: The inventory in BIN_1 is enough to satisfy the demand from class 1 in the period $(\tau_l, t+L]$, but the inventory in BIN_2 is not enough to satisfy all the demand from class 2. Therefore, there is inventory left over in BIN_1, but penalty cost is incurred for some of the class 2 demand arrivals. The units left in BIN_1 after satisfying all the demand from class 1 is l_1 and the holding cost is hl_1. Since there are some backorders for class 2, they incur delay cost of $p_2|l_2|$ and stock-out cost of $\lambda_2 \pi_2$. Note that l_2 is non-positive in this case. The expected penalty and holding cost per unit time for this case is

$$G_2(Q, S_1, S_2) = \sum_{l_1=1}^{\infty} \sum_{l_2=-\infty}^{0} (hl_1 + p_2|l_2| + \lambda_2 \pi_2) \times \\ g(S_1 - l_1, S_2 - l_2). \quad (8)$$

Case 3a: In this case, the inventory available in BIN_1 is not enough to satisfy the demand from class 1 in the period $(\tau_l, t+L]$ but the inventory in BIN_2 is enough to satisfy all the demand from class 2 during that period plus the demand from class 1 which is transferred over from BIN_1. Hence, there are no backorders for either class at time $t+L$. The cost for this case therefore is

$$G_{3a}(S_1, S_2, Q) = \sum_{l_1=-\infty}^{0} \sum_{l_2=|l_1|+1}^{\infty} h(l_1 + l_2) g(S_1 - l_1, S_2 - l_2). \quad (9)$$

As an example, suppose at time $t+L$, $l_1 = -5$ and $l_2 = 8$. This means that 5 units of class 1 cannot be

satisfied from BIN_1, but after satisfying all demand from class 2, BIN_2 still has 8 units left. So, the units left in BIN_2 can be used to fulfill the unsatisfied class 1 demand; then the units leftover will be $l_1 + l_2 = 3$, with the associated cost of $3h$.

Case 3b: In this case, the inventory available in BIN_2 after satisfying all the class 2 demand is not enough to satisfy the class 1 demand that spills over from BIN_1. Since there are l_2 units left in BIN_2 at least l_2 units of class 1 demand can be satisfied from BIN_2. However, the actual number of class 1 demand fulfilled from BIN_2 can be higher depending on the sequence of demand arrivals. When BIN_1 runs out of inventory, BIN_2 can be used to satisfy demand for both class 1 and class 2 on an FCFS basis, and therefore it may happen that some of the class 2 demand gets backordered as the BIN_2 gets exhausted by the class 1 demand arrivals, before some of the remaining class 2 demands arrive. We need to determine the possible sequences of demand arrivals and the probability of that sequence. For every such sequence, we can find the number of backorders for each class and the associated penalty cost, and, in turn, the expected penalty cost for this case. This case is best illustrated with an example which is discussed in the Appendix A. The expected holding and penalty cost for case 3b can be expressed as

$$G_{3b}(Q, S_1, S_2) = \sum_{l_1=-\infty}^{0} \sum_{l_2=1}^{|l_1|} \frac{\sum_{j=0}^{\min(S_2-l_2, |l_1+l_2|)} {}^{S_1+S_2}C_{S_2-l_2-j} \, {}^{|l_1+l_2|}C_j \left\{ \begin{array}{c} (|l_1 + l_2| - j)p_1 + \\ jp_2 + \lambda_1 \pi_1 + \lambda_2 \pi_2 \end{array} \right\}}{{}^{S_1+S_2-l_1-l_2}C_{S_2-l_2}} g(S_1 - l_1, S_2 - l_2). \quad (10)$$

Of the total demand $(S_1 + S_2 - l_1 - l_2)$ only $(S_1 + S_2)$ units of demand can be satisfied and the remaining $|l_1 + l_2|$ units of demand will be backordered. Since class 1 demand can also use BIN_2 after finishing BIN_1, demand from both classes is satisfied on an FCFS basis and only the last $|l_1 + l_2|$ demands will be backordered. Let j be the units of class 2 demand that arrives within the last $|l_1 + l_2|$ demand arrivals. Then, out of first $S_1 + S_2$ demands,

$S_2 - l_2 - j$ units would come from class 2. The probability of this event is $\frac{^{S_1+S_2}C_{S_2-l_2-j}{}^{|l_1+l_2|}C_j}{^{S_1+S_2-l_1-l_2}C_{S_2-l_2}}$. The units of class 1 demand within last $|l_1 + l_2|$ demands will be $(|l_1 + l_2| - j)$. So, the penalty cost is $\{(|l_1 + l_2| - j)p_1 + jp_2 + \lambda_1\pi_1 + \lambda_2\pi_2\}$. However, the total class 2 demand arrivals is $S_2 - l_2$ and hence j cannot be more than either $(S_2 - l_2)$ or $|l_1 + l_2|$. This leads to the expression given in Equation (10).

Case 4: In this case, the demand for both the classes in the period $(\tau_l, t+L]$ is greater than the allocated inventory position in the respective bins at time τ_l, that is, $D_1 \geq S_1$ and $D_2 \geq S_2$. While some of the class 1 demand can possibly be satisfied from BIN$_2$, as in case 3b, there are likely to be backorders for each class. The exact number of backorders for each class would again depend on the sequence of demand arrivals. The cost for case 4,

in the last $|l_1 + l_2|$ demand arrivals. Hence, $j > D_2 - (S_1 + S_2) = (-S_1 - l_2)$. This has to be reflected in the summation range in both terms in the numerator in Equation (11). Note that when $D_2 \leq (S_1 + S_2)$, $(-S_1 - l_2) \leq 0$, and then it does not have an impact on the summation range in either term. Finally, using Equations (7) to (11), the exact expression for the expected total cost of the proposed 2B policy can be obtained as

$$
\begin{aligned}
C(Q, S_1, S_2) = {} & \frac{\lambda A}{Q} + G_1(Q, S_1, S_2) + G_2(Q, S_1, S_2) \\
& + G_{3a}(Q, S_1, S_2) + G_{3b}(Q, S_1, S_2) + G_4(Q, S_1, S_2)
\end{aligned}
\tag{12}
$$

Note that, if we choose to optimize only the total (ordering + holding) cost subject to fill rate restric-

$$
G_4(Q, S_1, S_2) = \sum_{l_1=-\infty}^{0} \sum_{l_2=-\infty}^{0} \frac{\left(\begin{aligned} \sum_{j=\max(|l_2|+1,-S_1-l_2)}^{\min(S_2-l_2,|l_1+l_2|)} {}^{S_1+S_2}C_{S_2-l_2-j}{}^{|l_1+l_2|}C_j \left\{ \begin{aligned} & (|l_1+l_2|-j)p_1+ \\ & jp_2 + \lambda_1\pi_1 + \lambda_2\pi_2 \end{aligned} \right\} + \\ \sum_{j=\max(0,-S_1-l_2)}^{|l_2|} {}^{S_1+S_2}C_{S_2-l_2-j}{}^{|l_1+l_2|}C_j(|l_1|p_1 + |l_2|p_2 + \lambda_1\pi_1 + \lambda_2\pi_2) \end{aligned} \right)}{{}^{S_1+S_2-l_1-l_2}C_{S_2-l_2}} g(S_1 - l_1, S_2 - l_2).
\tag{11}
$$

As can be seen, the denominator in Equation (11) is the same as in Equation (10) in Case 3b. The first term of the numerator in Equation (11) is almost the same as in Equation (10). The major difference is that in Case 4, the demand for class 2 will incur stock out of at least $|l_2|$ units as its demand cannot be fulfilled from BIN$_1$ and the excess demand over the base stock level is $|l_2|$. When the number of class 2 demands, j, in the last $|l_1 + l_2|$ demands is less than $|l_2|$, it simply means that BIN$_2$ is completely exhausted by class 2 demands in the first $(S_1 + S_2)$ demands itself, and there is still stock available in BIN$_1$ to satisfy some of the class 1 demands in the last $|l_1 + l_2|$ demands in the period $(\tau_l, t+L]$. Therefore, expression Equation (11) has two terms, one for the case where the number of class 2 demands, in the last $|l_1 + l_2|$ demands, $j > |l_2|$ and one for the case when $j \leq |l_2|$. For $j \leq |l_2|$, the penalty cost incurred is identical for all values of j, and the stock out incurred is $|l_1|$ and $|l_2|$, respectively, for class 1 and class 2 demands.

The only other point to note in expression Equation (11) is that if $D_2 > (S_1 + S_2)$, then at least $D_2 - (S_1 + S_2)$ units of class 2 demand have to occur within the last $|l_1 + l_2|$ demand arrivals. Since a maximum of $(S_1 + S_2)$ units of class 2 demand can arrive within first $(S_1 + S_2)$ demand arrivals, the remaining units, $D_2 - (S_1 + S_2)$ can only arrive

tions on the two demand classes (as in Arslan et al. (2007)), the cost expression in Equations (10), (11), and (12) would be simplified significantly as $G_{3b} = G_4 = 0$, in that case.

2.2. Development of Optimal Policy Parameters

We now develop a method to derive the optimal policy parameters Q, S_1, and S_2 that minimize the total expected cost for the proposed 2B policy. For the CLR policy, Deshpande et al. (2003) showed that for a fixed value of the critical inventory level K, Q can be determined using an algorithm similar to Federgruen and Zheng (1992), by exploiting the convexity of the underlying cost terms. However, they still needed to enumerate the cost exhaustively for all possible values of K.

Due to the complex nature of the cost expression of the proposed 2B policy, no such structural results on the cost function could be obtained. However, we are able to use the key insights from the (Q, r) policy algorithm of Federgruen and Zheng (1992) to limit our search in the algorithm to determine the optimal policy parameters. Note that in our case, $S_1 + S_2 = r + Q$. Let $S_1 * (Q)$ and $S_2 * (Q)$ be the optimal value of S_1 and S_2 for a given Q. In the Federgruen and Zheng algorithm for the single demand class problem, when Q is incremented by one to $Q_{new} = Q + 1$, either the

optimal value of r remains the same, or r decreases by 1 [or in other words, the new value of $(r + Q)$ either remains the same (r decreases by 1, Q increases by 1) or increases by 1 (r remains same, Q increases by 1)]. By the same logic, in our proposed 2B policy, when Q is incremented by 1 to $Q_{new} = Q + 1$, there are three possible values for $S_1*(Q + 1)$, $S_2*(Q + 1)$. The possible values are (1) $S_1*(Q + 1) = S_1*(Q)$ & $S_2*(Q + 1) = S_2*(Q)$, (2) $S_1*(Q + 1) = S_1*(Q) + 1$ & $S_2*(Q + 1) = S_2*(Q)$, and (3) $S_1*(Q + 1) = S_1*(Q)$ & $S_2*(Q + 1) = S_2*(Q) + 1$. In other words, when Q is incremented in the algorithm, we need only search for three possible values of (S_1, S_2). While we are not able to prove this result analytically, all our numerical experiments over a range of problem parameters support this conjecture. This, along with the bounds on the decisions variables, Q and $\overline{S} = S_1 + S_2$ (developed below), help us to search for the optimal values for the proposed policy in an efficient manner.

2.3. Bounds on $\overline{S} = S_1 + S_2$ and Q
Clearly a lower bound, $\overline{S_L}$, on $\overline{S}$ can be obtained by solving a single customer class problem with demand, $\lambda = \sum_{i=1}^{2}\lambda_i, \pi = \pi_2$ and $p = p_2$. An upper bound, $\overline{S_{U1}}$, on $\overline{S}$ can be obtained by solving a single customer class problem with $\lambda = \sum_{i=1}^{2}\lambda_i, \pi = \pi_1$ and $p = p_1$.

If a separate stock policy was considered (i.e., BIN_2 can only be used to satisfy demand from class 2, but ordering for the two classes is done together), then the problem reduces to a joint replenishment problem (JRP) and the (Q, S) policy developed by Pantumsinchai (1992) can be used to solve it. As the inventory of each BIN_i is segregated to fulfill only a specific demand class, the value of $S_1 + S_2$ of this JRP solution should serve as an upper bound $\overline{S_{U2}}$ on $\overline{S}$ of the original problem. One can then choose the upper bound on $\overline{S}$ as $\overline{S_U} = \min(\overline{S_{U1}}, \overline{S_{U2}})$. If each demand class were to be treated as a separate item, each with ordering cost A, and demand λ_i (i.e., a separate stock policy with no economies of scale in ordering), then the sum of the optimal order quantity of the two separate classes or items would clearly be an upper bound, Q_U, on the Q.

The algorithm for determining the optimal policy then involves searching for the lowest C (Q, S_1, S_2) within the lower and upper bounds of Q and $\overline{S}$. For the first value of Q, we have to search for all values of (S_1, S_2) within the lower and upper bounds of $\overline{S}$. Thereafter, when Q is incremented by one in each iteration, one needs to search only for the three possible combinations of (S_1, S_2) as mentioned earlier.

3. Numerical Results

The performance of the proposed 2B policy was compared with the CLR policy in an extensive numerical study. The problem parameters used in the study were similar to that used in Deshpande et al. (2003), which itself reflect data from different categories of industries. In the first part of the numerical study, we included the penalty cost in the total cost and did not impose any minimum service level requirements on the two demand classes. The problem parameters used in the first numerical study are summarized in Table 3. We initially had 840 problems in the first numerical study. However, for values of $p_2 > 1200$, both the policies resulted in identical optimal cost and with no need for rationing for a significant majority of the problems. Hence only two values of p_2 were considered and this reduced the number of problems in the study to 168. For all these 168 problems, we calculated the optimal cost (as well as the fill rates obtained for the two demand classes) for both the policies.

The service level for the two classes is determined using PASTA property which states that for Poison demand, the fill rate is equal to the probability that the steady-state inventory level is positive. For the CLR policy, the fill rates for the two demand classes can be calculated as

$$SL_1^{CLR} = \Pr(IL > 0), \qquad (13)$$

$$SL_2^{CLR} = \Pr(IL \geq K), \qquad (14)$$

where, K is the rationing level and IL is the steady-state inventory level. For the proposed 2B policy, the fill rates for the two demand classes can be calculated as follows:

$$SL_1^{2B} = \Pr(l_1 > 0) + \Pr[l_1 \leq 0 \ \& \ ((l_1 + l_2) > 0)], \quad (15)$$

$$SL_2^{2B} = \Pr(l_2 > 0). \qquad (16)$$

3.1. Comparison of The Two Policies without Service Level Constraints
The complete results for the two policies for the first numerical study can be obtained from the authors.

Table 3 Problem Parameters Used in the First Numerical Study

Demand	$\lambda = (\lambda_1 + \lambda_2) = 20$, $\lambda_1 = 7, 8, 9, 10, 11, 12, 13$ (7 values)
Back order cost	Delay cost $p_1 = 6000$ (fixed). $p_2 = 600, 1200$ (2 values)
	Stock-out cost $\pi_1 = \pi_2 = 0$
Lead time	$L = 0.25, 0.3, 0.35, 0.4, 0.45$, and 0.5 (6 values)
Holding cost rate	$h = 250, 300$ (2 values)
Fixed ordering cost	$A = 100$

Table 4. **Comparison of The 2B Policy with CLR Policy for Problems Without Service Level Constraints: Partial Results**
($\lambda = (\lambda_1 + \lambda_2) = 20$, $p_1 = 6000$, $p_2 = 600$, $\pi_1 = \pi_2 = 0$, $h = 250$, $A = 100$)

		Performance difference between 2B policy and CLR policy		
L	λ_1 ($\lambda_2 = 20 - \lambda_1$)	Percentage cost difference between the 2B policy and the CLR policy (%)	Difference in fill rate for the higher class demand (SL_1) between the 2B policy and the CLR policy (%)	Difference in fill rate for the lower class demand (SL_2) between the 2B policy and the CLR policy (%)
0.25	7	5.03	2.20	4.90
	8	5.48	5.70	11.70
	9	4.60	6.20	9.30
	10	4.29	5.60	22.80
	11	3.87	6.00	20.10
	12	3.79	0.70	14.70
	13	3.63	2.00	16.60
0.3	7	5.89	6.00	11.00
	8	5.08	6.60	8.60
	9	4.87	7.10	24.80
	10	4.27	7.30	23.40
	11	4.31	2.90	14.10
	12	4.37	0.60	12.40
	13	4.33	0.80	10.40

A partial sample of the results is provided in Table 4. For all the 168 problems in the study, the CLR policy resulted in a lower cost than the 2B policy. The average difference in the cost between 2B and CLR policy was 3.85%, and the maximum difference was 6.83%. However, the service level provided to the lower class demand by the proposed 2B policy was as much as 28% higher than that of the CLR (with an average increase in fill rate of 11.5%). The 2B policy also resulted in a slightly higher fill rate for the higher class demand, with an average increase in fill rate of 2.8%, and a maximum increase of 9%. Thus, the 2B policy provides a much higher service level to the lower class demand, for only a slightly increase in cost (less than 4%). This is because the inventory in BIN_2 initially gets reserved for the lower class demand as the demand for the higher class gets fulfilled first from BIN_1, whereas in CLR, demand from both the classes are initially fulfilled from the same common stock.

3.2. Comparison of The Two Policies with Service Level Constraints

When no service level constraints are imposed and the total cost (including penalty cost), is optimized, the CLR policy provides a low fill rate for the lower priority demand class (which is as low as 64%). In reality, the service level requirement for the lower demand class may not be that low.

We therefore considered a service level constrained optimization problem (as in Arslan et al. 2007), in the second part of the numerical study. In this case, the total cost (comprised of ordering cost and inventory holding cost) is optimized subject to minimum fill rate constraints on both the demand classes. The optimal solution for both the 2B and CLR policies were obtained by exhaustively searching for the lowest cost solution that satisfied the fill rate requirements for the two demand classes.

The problem parameters used for the experiments in the second part of the numerical study were the

Table 5 **Comparison of The 2B Policy with CLR Policy for The Service Level Constrained Optimization Problem**

SL_1	99%	99%	99%	99%	95%	95%	95%	90%	90%	85%	85%	All problems
SL_2	95%	90%	85%	80%	90%	85%	80%	85%	80%	80%	75%	
Cost difference between CLR and 2B policy ((CLR-2B)/CLR) in percentage												
Minimum difference	−8	−13	−13.1	−18.1	−4.7	−11.3	−11.1	0	−6.2	0	−6.2	−18.1
Maximum difference	1	−1.2	−3.9	−1.8	4.5	−0.1	−0.6	5.2	1.8	5.8	5.2	5.8
Average difference	−2.8	−7.5	−9	−10.4	−0.3	−4.6	−5	2.4	−2.7	2.1	−0.4	−3.5
Percentage of problems where CLR had lower cost	90	100	100	100	39	100	100	0	86	0	38	69
Percentage of problems where 2B had lower cost	8	0	0	0	61	0	0	60	8	43	43	20
Percentage of problems where both policies had identical cost	1	0	0	0	0	0	0	40	6	57	19	11

SL_1 is the minimum fill rate required for demand class 1 and SL_2 is the minimum fill rate required for demand class 2.

10

same as that given in Table 3, but with no penalty costs for backordering. Instead, different values of the fill rate constraints for the two demand classes were used in this study. For each set of minimum fill rate combinations for the two demand classes (SL$_1$, SL$_2$) there were a total of 84 problems (7 demand values, 6 values of the lead time, and 2 values for the holding cost rates). A total of 11 different fill rate combinations were used in the study ([99%, 95%], [99%, 90%], [99%, 85%], [99%, 80%], [95%, 90%], [95%, 85%], [95%, 80%], [90%, 85%], [90%, 80%], [85%, 80%], and [85%, 75%]). Therefore a total of 924 problems were solved in the second part of the numerical study.

The complete numerical results for these set of experiments can also be obtained from the authors. Table 5 provides a summary of these results. The results of the numerical study show that CLR dominates 2B in 69% of the problems, and 2B dominates CLR in 20% of the problems. 2B seems to dominate in a large percentage of the problems when the service level difference for the two classes is not too high, or when the service levels required are relatively lower. For example, when (SL$_1$, SL$_2$) = (90%, 85%), the 2B policy dominated the CLR policy in 60% of the problems and for the remaining 40% of the problems, both policies resulted in the same optimal cost. By its very nature of keeping stocks in two separate bins, the 2B policy works well when the service level differentiation required between the two demand classes is relatively small. The CLR policy on the other hand works well when a larger difference in the service level is required.

4. Summary

In this study, a new class of two bin (2B) policy is developed for continuous review inventory systems with two differentiated demand classes. We have developed the cost expressions for the proposed policy under the assumption of clearing the backorders using a threshold clearing mechanism. An algorithm to determine the optimal policy under this class is also proposed. When the total cost (including penalty cost) is optimized without any fill rate restrictions, the CLR policy dominates 2B policy in all problems in a computational study. However, the proposed 2B policy provides a higher service level (especially for the lower customer class) compared to the CLR policy, without too much increase in the cost. The fill rate provided for the lower priority class in the CLR policy can be very low (as low as 64%). Instead of penalty costs, if minimum fill rate restrictions are imposed on the two demand classes, the proposed 2B policy outperforms the CLR policy, when the difference in service level requirement for the two demand classes is not very high. Future research could possibly address

extending the proposed 2B policy to systems involving more than two customer classes and to other demand distributions.

Acknowledgments

Appendix
Numerical Example Illustrating Case 3b of the Policy Cost Calculation

Let $S_1 = 7$ and $S_2 = 6$, $l_1 = -3$ and $l_2 = 1$. The demand for the two classes during the period $(\tau_l, t + L]$ is $(S_1 - l_1) = 10$, and $(S_2 - l_2) = 5$, respectively, with the total demand being 15. Of the total demand of 15 for both classes together, a maximum of only $S_1 + S_2 = 13$, can be satisfied. Of the 10 units of demand from class 1, 7 would be satisfied from BIN$_1$. The exact number of units of class 1 demand (in excess of 7) that would be satisfied depends upon the sequence of arrival of the class 1 and class 2 demands. There are three possible ways these demands can arrive.

(a) Of the first 13 units of demand arrivals, 10 units are from class 1 and 3 units are from class 2 (maximum possible demand from class 1 is 10) and the last 2 (14th and 15th) demand arrivals are from class 2.

(b) Of the first 13 units of demand arrivals, 9 units are from class 1 and 4 units are from class 2. Of the last two demand arrivals, 1 is from class 1 and the other from class 2.

(c) Of the first 13 units of demand arrivals, 8 units are from class 1 and 5 units are from class 2 and the last 2 demand arrivals are from class 1.

No other case is possible since demand from class 2 cannot exceed 5. In the first case, all the class 1 demand can be satisfied and there will be no backorders for class 1. Also 3 units of class 2 demands will be satisfied and 2 units of demand which arrived last will be backordered. In the second case, 1 unit of the demand for each class will be backordered. Finally, in the third case, all the class 2 demands are satisfied and 2 units of class 1 will be backordered as it arrives after BIN$_2$ becomes empty.

The probability of the first case is $\frac{^{13}C_3 {}^2C_2}{^{15}C_5}$. The numerator denotes the number of ways in which there can be 3 class 2 demands out of first 13 demands and 2 class 2 demands out of last 2 demands. The denominator denotes the number of ways in which there can

be 5 class 2 demands of total 15 demands. Similarly, the probability for the second case and third case are $\frac{^{13}C_4 {}^2C_1}{^{15}C_5}$ and $\frac{^{13}C_5 {}^2C_0}{^{15}C_5}$, respectively. Since in the first case, there are no class 1 backorders and 2 units of class 2 backorders, the penalty cost is $(0p_1 + 2p_2 + \lambda_1\pi_1 + \lambda_2\pi_2)$. Similarly, for the second and third cases it will be $(1p_1 + 1p_2 + \lambda_1\pi_1 + \lambda_2\pi_2)$ and $(2p_1 + 0p_2 + \lambda_1\pi_1 + \lambda_2\pi_2)$, respectively. So, the expected penalty cost is $\frac{^{13}C_3 {}^2C_2}{^{15}C_5} (0p_1 + 2p_2 + \lambda_2\pi_2) + \frac{^{13}C_4 {}^2C_1}{^{15}C_5} (1p_1 + 1p_2 + \lambda_1\pi_1 + \lambda_2\pi_2) + \frac{^{13}C_5 {}^2C_0}{^{15}C_5} (2p_1 + 0p_2 + \lambda_1\pi_1 + \lambda_2\pi_2)$. Note that due to PASTA property and the fact that there is no inventory left over in either bins at time $(t + L)$, the expected stock-out cost for demand class k is $\lambda_k\pi_k$.

References

Alvarez, E. M., M. C. Heijden, W. H. M. Zijm. 2013. Service differentiation in spare parts supply through dedicated stocks. *Ann. Oper. Res.*, doi 10.1007/s10479-013-1362-z.

Arslan, H., S. C. Graves, T. Roemer. 2007. A single-product inventory model for multiple demand classes. *Manage. Sci.* **53**(9): 1486–1500.

Dekker, R., R. M. Hill, M. J. Kleijn. 2002. On the (S -1, S) lost sales inventory model with priority demand classes. *Nav. Res. Logistics* **49**(6): 593–610.

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

Evans, R. V. 1968. Sales and restocking policies in a single inventory system. *Manage. Sci.* **14**(7): 463–473.

Federgruen, A., Y. Zheng. 1992. An efficient algorithm for computing an optimal (r, Q) policy in continuous review stochastic inventory systems. *Oper. Res.* **40**(4): 808–813.

Gayon, F., F. de Véricourt, F. Karaesmen. 2009. Stock rationing in a multi-class make-to-stock queue with information on the production status. *IIE Trans.* **41**(12): 1096–1109.

Ha, A. Y. 1997a. Inventory rationing in a make-to-stock production system with several demand classes and lost sales. *Manage. Sci.* **43**(8): 1093–1103.

Ha, A. Y. 1997b. Stock rationing policy for a make-to-stock production system with two priority classes and backordering. *Nav. Res. Logistics* **44**(5): 457–472.

Ha, A. Y. 2000. Stock rationing in an M/E$_k$/1 make-to-stock queue. *Manage. Sci.* **46**(1): 77–87.

Kaplan, A. 1969. Stock rationing. *Manage. Sci.* **15**(5): 260–267.

Melchiors, P., R. Dekker, M. J. Kleijn. 2000. Inventory rationing in an (s, q) inventory model with lost sales and two demand classes. *J. Oper. Res. Soc.* **51**(1): 111–122.

Moon, I., S. Kang. 1998. Rationing policies for some inventory systems. *J. Oper. Res. Soc.* **49**(5): 509–518.

Nahmias, S., S. Demmy. 1981. Operating characteristics of an inventory system with rationing. *Manage. Sci.* **27**(11): 1236–1245.

Pantumsinchai, P. 1992. A comparison of three joint ordering inventory policies. *Decision Sci.* **23**(1): 111–127.

Topkis, D. 1968. Optimal ordering and rationing policies in a non-stationary dynamic inventory model with N demand classes. *Manage. Sci.* **15**(3): 160–176.

Veinott, A. F. 1965. Optimal policy in a dynamic, single product, non stationary inventory model with several demand classes. *Oper. Res.* **13**(5): 761–778.

de Véricourt, F., F. Karaesmen, Y. Dallery. 2000. Dynamic scheduling in a make-to-stock system: A partial characterization of optimal policies. *Oper. Res.* **48**(5): 811–819.

de Véricourt, F., F. Karaesmen, Y. Dallery. 2002. Optimal stock allocation for a capacitated supply system. *Manage. Sci.* **48**(11): 1486–1501.