PERGAMON

Computers & Operations Research 28 (2001) 935–953

computers & operations research

www.elsevier.com/locate/dsw

A dynamic two-segment partial backorder control of (r, Q) inventory system

C.W. Chu[a], B.E. Patuwo[a], A. Mehrez[b], G. Rabinowitz[b,]*

[a]Graduate School of Management, Kent State University, Kent, OH 44242, USA

[b]Department of Industrial Engineering and Management, Ben Gurion University of the Negev, Beer-Sheva, 84105, Israel

Received 1 May 1999; received in revised form 1 December 1999

Abstract

Inventory management involves determination of shortage policy. It specifies the conditions for losing or backordering a demand. Alternative policies include pure backorder, pure lost sales, and partial backorder (using a single backorder control limit). When the backorder-cost is time dependent it makes sense to modify the backorder-limit over time. Thus, a new form of partial backorder policy (PB2) with two-segment backorder control limits is introduced. The traditional policies mentioned above, are special cases of PB2. Hence, we provide a unified framework for studying different policies that deal with shortage. The PB2 problem is formulated and solved as a discrete time, stochastic constrained control problem. Its performance is numerically compared with the simpler alternative policies. In some cases its cost savings, versus the best of PB and PL, exceeds 15%, and 7% versus a single backorder limit policy. The economical advantage is significant over a wide range of the problem parameters. © 2001 Elsevier Science Ltd. All rights reserved.

Scope and purpose

This paper develops an expanded framework for modeling shortages in inventory management. It recognizes that optimal backordering strategy may change over time during an "out-of-stock" period. The paper is motivated by experience in the chemical industry in which, the cost of backordering is highly time related. Inventory managers, in this industry, consider to lose sales initially (once they run out of stock) and begin to backorder demand later as they approach the replenishment time. A two-segment partial backorder (r, Q) model is introduced and solved. Pure backorder (PB), pure lost sales (PL), and partial backorder (using a single backorder limit), are all special cases of the proposed model. The problem is formulated and solved as a discrete time, stochastic constrained control problem. Its performance is numerically compared with the simpler alternative policies. In some cases its cost savings, versus the best of PB and PL, exceeds 15%, and

* Corresponding author. Tel.: + 972-7-647-2199; fax: + 972-7-647-2958.
E-mail address: rgadi@bgumail.bgu.ac.il (G. Rabinowitz).

PII: S0305-0548(00)00017-4

7% versus a single backorder limit policy. The economical advantage is significant over a wide range of the problem parameters. The partial backorder policy we propose is not only different from those in the literature, but it provides new control flexibility.

Keywords: Inventory; Continuous review; Partial backorder; Lost sales; Backorder; Stochastic demand

1. Introduction

When a firm is a sole supplier, or when there is a lack of substitutions or competitors, customers may prefer to backorder unfilled demands. Lost sales occur when the customers prefer not to wait for the next replenishment, or when the firm decides to buy similar items from competitors to satisfy demands and maintain customer loyalty. However, as pointed out by Peterson and Silver [1, p. 253], in most practical situations, one finds a combination of these two extreme policies where some of the excess demands are backordered and the rest are lost.

Additional motivation for this study stems from consulting experience with inventory control systems in the chemical industry. There, the partial backorder (PB) policy is implemented in the following manner: out-of-stock items are first backordered (because there is no a priori knowledge of how many items will be short during the lead time). When the quantity of the backordered items reaches a certain limit, similar items are purchased from competitors to satisfy demands. We found that some companies would rather do this than lose market share. A model for this type of partial backorder policy is implemented using an explicit control limit b, which is the maximum allowable number of backorders during the lead time. Real-time control implementation of this policy is easy; only the backordered quantities are to be tracked. This partial backorder policy is called PB1 policy and was analyzed by Rabinowitz et al. [2].

In this paper, a more general policy called PB2 is introduced. It extends PB1 by allowing two time segments during the lead time, to enable two backorder control limits b_1 and b_2. It provides another instrumental control to further reduce cost without adding complexity to the real-time implementation. Formal descriptions of PB1 and PB2 policies are given in Section 2. There, we will demonstrate that pure backorder, pure lost sales and PB1 policies are special cases of PB2. Any policy which backorders during the first segment of the lead time and lose sales during the second, or inversely, is also a special case of PB2. Therefore, the PB2 policy provides a unified framework for studying different policies, which deal with shortage. The cost function of PB2 is given in Section 3 and the solution procedure is provided in Section 4. A numerical investigation in Section 5 explores the sensitivity of the solution to the unit shortage costs (backorder and lost sales). We demonstrate that the PB2 policy is better than the other three policies for a certain range of cost parameters. In some conditions, cost savings of more than 10% are realized. Potential future research is discussed in the last section.

The literature, on partial backorder inventory control is sparse. An identical policy to the one we propose was not found. Montgomery et al. [3] consider a continuous review inventory system where a fraction α, of the unfilled demand is backordered and the remainder is lost. Almost all articles in the literature model partial backorder in a similar way, namely, not as a decision variable. Cases of both deterministic and stochastic demands are considered, but the stochastic demand case is treated heuristically. Rosenberg [4] reformulates the model of Montgomery et al. [3] by introducing a "fictitious demand rate" which simplifies the analysis of a partial backorder policy and gives an economic interpretation to the circumstances under which this policy is optimal.

As in our case, Kim and Park [5] consider a continuous review system with a constant lead time where the backorder cost is assumed to be proportional to the length of time for which the backorder exists. However, their backorder control variable is different, it is a fraction of the unfilled demand. Assuming at most one outstanding order at any point in time, they derive the equations from which r (the reorder point) and Q (the order size) can be computed iteratively.

The literature proposes a variety of ways to deal with backordering. Under Poisson demand and exponential lead-time, Woo and Sphicas [6] formulate a partial backorder model, which allows for a finite number of orders to be outstanding. Steady-state probabilities for the inventory level are first derived and then, optimal values of r and Q are found using a search method. In our case, we assume a single outstanding order and provide a closed-form solution for Q, for any combination of the remaining decision variables, which are solved by search.

Posner [7] treats the case where backordered customers are willing to wait for a random duration. In this model the partial backorder can be considered as a performance measure but is not a decision variable. Das [8] employs an $(S-1, S)$ policy and assumes Poisson demand with a constant time limit on the backordered demand. Moinzadeh [9] extends it to allow partial backorders. Smeitink [10] proves Moinzadeh's results, that the steady-state net inventory probabilities depend on the mean of the lead time and not on the shape of its distribution.

Porteus [11] reviews periodic review models including one where a fraction of the excess demands is backordered. A myopic approximation to this model is provided by Nahmias [12]. For recent findings regarding the computations of optimal solutions to general (s, S) inventory systems with backorder policy (both periodic review and continuous review systems), see Zheng and Federgruen [13,14]. For continuous review backorder systems, see Federgruen and Zheng [15] and for the discussions of the sensitivity of the optimal solutions, see Zheng [16]. The partial backorder policy we propose is not only different from those in the literature, but it provides new control flexibility with significant economical advantage.

2. Partial backorder policies

For the convenience of the reader, before defining the PB2 policy, we will first define a simpler partial backorder policy called PB1, which was analyzed in Rabinowitz et al. [2]. Later on we show that PB1 is a special case of PB2. The demand process is assumed to be Poisson with rate λ. We will use the following standard notations of the continuous review system: r is the reorder point; Q is the order quantity; τ is the fixed lead time; $I(t)$ is the inventory level at time t, where the inventory level is defined as the stock on-hand minus the number of backorders. When $I(t) < 0$, we have a shortage, with $-I(t)$ the amount backordered. The cycle time T, is defined as the time between reorders.

2.1. PB1 policy

Let b (a nonnegative integer) denote *the backorder limit*. For $0 < t < \tau$, if $-b+1 \leqslant I(t) \leqslant 0$, a demand that arrives at time t is backordered, and if $I(t) = -b$, that demand is considered lost. In other words, if $D(\tau)$ is the demand during the lead time, the first r units demanded are satisfied from stock, the next b units demanded are backordered, and the remaining are lost, thus, b is also the maximum number of units backordered. If $b = 0$, we have the pure lost sales policy, and if b is sufficiently large so that no lost sales are possible, we have the pure backorder policy. Fig. 1a, which is a realization of inventory level against time, illustrates the PB1 policy.

2.2. PB2 policy

Under PB1, a single-decision variable (b) determines the backorder policy. Under PB2 we define three decision variables for this purpose as follows. First, t_1 (denoted *the intermediate review time*) a selected time interval, starting at reorder and ending no later than the replenishment: $0 \leqslant t_1 \leqslant \tau$. Second, b_1 and b_2 (nonnegative integers) *the backorder limits*, with $b_1 \leqslant b_2$. The lead-time is divided by t_1 into two time segments. The maximum number of units backordered is b_1 within the first time segment and b_2 within the second one. An arriving demand at time t during the first time segment, is backordered if $-b_1+1 \leqslant I(t) \leqslant 0$, but is lost if $I(t) = -b_1$. During the second time segment, an arriving demand is backordered if $-b_2+1 \leqslant I(t) \leqslant 0$, and is lost if $I(t) = -b_2$. Fig. 1b illustrates the PB2 policy.

If $b_1 = b_2$, then PB2 reduces to PB1. If $b_1 = b_2 = 0$, it reduces to the pure lost sales policy and if b_1 is sufficiently large, it becomes the pure backorder policy. If $b_1 = 0$ and $b_2 > 0$ the PB2 policy denies backorder in the first time segment but permits a partial backorder during the second segment. When backorder cost includes a time-dependent factor, such a policy makes sense, but is

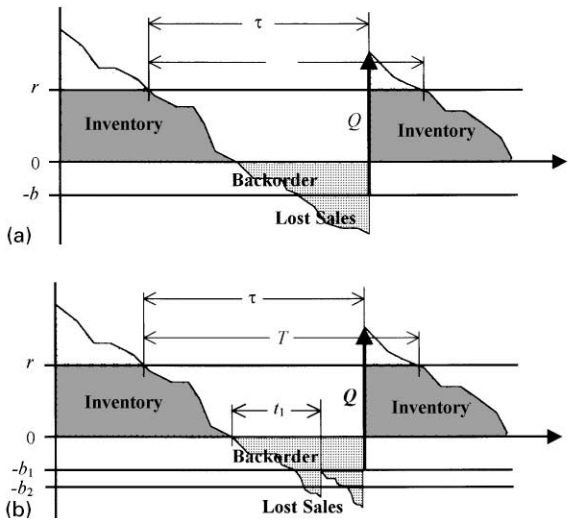


Fig. 1. (a) The inventory over time under PB1 policy. (b) The inventory over time under PB2 policy.

not enabled under any previous model, including PB1. Apparently, the Two-segment PB2 policy can be extended to include any finite number of time segments of *backorder limits* within the lead time.

The *backorder limits*, b_1 and b_2, the first time-segment duration t_1, the reorder point r and the order quantity Q, are the model decision variables. We would like to identify the conditions, if exist, for which an optimal PB2 policy is different than that of PB1, pure backorder, and pure lost sales policies. If such a situation exists, we need to examine its economical advantage versus the other three policies.

The cost function will be derived in the next section and the formulas are obtained by employing results from regenerative processes (see, for example, Cinlar [17]) and by following Hadley and Whitin's derivation [18, pp. 197–200] of the lost sales case. In Section 4, we show that when r, t_1, b_1 and b_2 are fixed, the optimal order quantity Q^* can be found analytically. The variable t_1 is continuous, but the first partial derivative of the expected cost function with respect to t_1 is too complex, thus we employ a crude discrete approximation to solve for t_1. The discrete approximation of t_1 is justified because its resolution can be freely chosen, to obtain any desired accuracy. In addition, an inventory control switching time, in practice, is not continuous, but is selected from a discrete time scale (e.g. days). Consequently, a four-dimensional discrete search is used to identify the optimal r, b_1, b_2 and t_1.

3. The cost function

For ease of exposition, we use a *year* to denote the time unit. Following Hadley and Whitin [18], we incorporated three unit shortage costs:

π = per unit cost of lost demands,
$\hat{\pi}$ = per unit cost of backorders,
π' = per unit-year cost of backordered demands,

and the ordinary unit costs:

h = unit holding cost,
K = fixed ordering cost,
c = variable unit purchasing cost.

We further assume that the demand follows a stationary Poisson process with a rate of λ, a demand that arrives at time t is satisfied immediately as long as the inventory level $I(t) > 0$, and an order of size Q is placed when $I(t) = r$ and is received τ years later (a deterministic lead-time). In addition, we bound the inventory control policy such that at most one order is outstanding, as in Hadley and Whitin [18, Sections 4–11]. Clearly, such a bound might reduce the model applicability especially when the lead time is long and the inventory holding cost is high. It is, however, a common bound both in theory and in practice. Under this bound, $\{I(t)\colon t \geqslant 0\}$ forms a regenerative process with renewals at the reorder times. This is the main motivation for this limitation, relaxing this bound would complicate the model significantly.

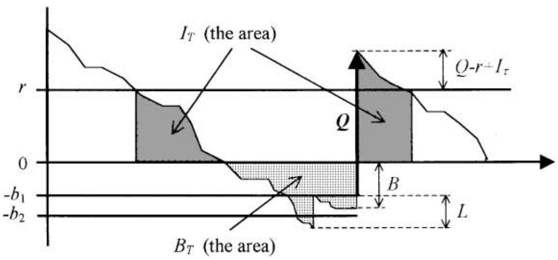


Fig. 2. The random state variables under PB2 policy.

Let T be the random cycle time, between two successive reorders. From renewal theory, the expected annual cost is given by

$$\bar{C}(r, b_1, b_2, t_1, Q) = \frac{E[C]}{E[T]}, \tag{1}$$

where C is the random total cost within a cycle. Here,

$$E[C] = K + cQ + hE[I_T] + \pi E[L] + \hat{\pi} E[B] + \pi' E[B_T], \tag{2}$$

where I_T is the accumulated (unit-year) inventory time within a cycle, L is the number (units) of lost sales per cycle, B is the number (units) of backorders in a cycle, and B_T is the accumulated (unit-year) backorder time during a cycle. Fig. 2 illustrates these random state variables. The expected values $E[I_T]$, $E[L]$, $E[B]$, $E[B_T]$, and $E[T]$ will be derived later.

In the following, we adopt the notations and identities used by Hadley and Whitin [18, Appendix 3]. Thus, we denote the Poisson probability as $p(j;\ \lambda t) = (\lambda t)^j e^{-\lambda t}/j!$ and its tail probability as $P(j;\ \lambda t) = \sum_{i=j}^{\infty} p(i;\ \lambda t)$, for $j = 0,\ 1, \ldots$.

3.1. Inventory level at τ

Note that the state space of the inventory level just before replenishment arrives, $I(\tau)$, is $\{-b_2, -b_2+1, \ldots, -b_1, \ldots, 0, \ldots, r-1, r\}$. Its kth moment is derived in the appendix by Eq. (A.1), based on the following decomposition:

$$E[I(\tau)^k] = E_{D_1}[E(I(\tau)^k | D_1)], \tag{3}$$

where D_1 is the demand during $(0, t_1]$ the first time segment of the replenishment period.

3.2. Expected backorder per cycle

The expected number of backorders B in a cycle, is derived by Eqs. (A.2) and (A.3) in the appendix, based on the following decomposition:

$$E[B] = E[B_1] + E_{D_1}[E(B_2 | D_1)], \tag{4}$$

where B_1 is the random number of backorders during $(0, t_1]$, and B_2 during $(t_1, \tau]$.

3.3. Expected backorder time per cycle

The expected value of the accumulated (unit-year) backorder time during a cycle, $E[B_T]$, is given by

$$E[B_T] = \int_0^{t_1} E[B_1(t)]\,\mathrm{d}t + \int_{t_1}^{\tau} E[B_2(t)]\,\mathrm{d}t, \tag{5}$$

where $B_1(t)$ is the number of backorders at time $t \in (0, t_1]$, and $B_2(t)$ is at $t \in (t_1, \tau]$. The derivation is provided in the appendix by Eqs. (A.4)–(A.6).

3.4. Expected inventory per cycle

The expected accumulated (unit-year) inventory time held per cycle $E[I_T]$ consists of two parts: *before* the replenishment, denoted by $E[I_B]$, and after the replenishment, denoted by $E[I_A]$. Denoting $D(t)$ for $0 \leqslant t \leqslant \tau$ as the demand during $(0, t]$:

$$E[I_B] = \int_0^{\tau} E[\max\{0,\, I(t)\}]\,\mathrm{d}t = \int_0^{\tau} E[\max\{0,\, r - D(t)\}]\,\mathrm{d}t$$

$$= \int_0^{\tau} \left[\sum_{x=0}^{r-1} (r - x)p(x;\, \lambda t)\right]\mathrm{d}t = r\sum_{x=0}^{r-1} g(x,\, \lambda,\, \tau) - \sum_{x=0}^{r-1} xg(x,\, \lambda,\, \tau), \tag{6}$$

where $g()$ is defined by Eq. (A.5) in the appendix.

The expected accumulated (unit-year) inventory time *after* the replenishment is derived by Eqs. (A.7)–(A.8) in the appendix, and is given by

$$E[I_A] = \frac{1}{2\lambda}\{E[I(\tau)^2] + 2QE[I(\tau)] + Q^2 - r(r+1), \tag{7}$$

where $E[I(\tau)^2]$ and $E[I(\tau)]$ are the first and second moments of $I(\tau)$ which are given by Eq. (A.1) in the appendix.

Summing up $E[I_A]$ and $E[I_B]$ we obtain

$$E[I_T] = r\sum_{x=0}^{r-1} g(x,\, \lambda,\, \tau) - \sum_{x=0}^{r-1} xg(x,\, \lambda,\, \tau)$$

$$+ \frac{1}{2\lambda}\left\{E[I(\tau)^2] + 2QE[I(\tau)] + Q^2 - r(r+1)\right\}. \tag{8}$$

3.5. The expected cycle time

The cycle time can be divided into two parts: the time from reorder to replenishment, which is the lead time τ and the time from replenishment to the next reorder time, denoted by T_A. By definition, the expected cycle length $E[T] = \tau + E[T_A]$. To find $E[T_A]$, we define $Z = I(\tau) + Q$ as the inventory level just after replenishment. The state space of Z is $\{-b_2 + Q, \ldots, r + Q\}$. Thus, $N = Z - r$ represents the demand during T_A. This means that T_A equals the arrival time of N units of demand: $T_A = \sum_{i=1}^{N} T_i$ where T_i's are iid exponentially distributed with parameter λ

representing the demand rate. Since N is a stopping time for the renewal process $\{T_1, T_2, \dots\}$ and $E[N] < \infty$, from Wald's equation (see Ross [19, p. 59]), we obtain:

$$E[T_A] = E\left[\sum_{i=1}^{N} T_i\right] = E[N]E[T_i] = \frac{1}{\lambda}E[N].$$

Since $N = Z - r = I(\tau) + Q - r$, with a state space: $\{-b_2 + Q - r, \dots, Q\}$, we obtain

$$E[N] = Q - r + E[I(\tau)],$$

where $E[I(\tau)]$ is given by Eq. (A.1) in the appendix. Therefore,

$$E[T] = \tau + E[T_A] = \tau + \frac{1}{\lambda}\left\{Q - r + E[I(\tau)]\right\}. \tag{9}$$

By setting $b_1 = b_2 = b$, one may confirm that $E[T]$ reduces to its structure under PB1 in Rabinowitz et al. [2].

3.6. Expected lost sales per cycle

Clearly, in the long run, if all demands during a cycle are to be supplied (no lost demand), then the expected cycle length would be Q/λ. Thus, including the possibility of lost demands,

$$E[T] = \frac{Q}{\lambda} + E[T_L], \tag{10}$$

where T_L is the portion of the cycle with lost demands. Using (9), we get

$$E[T_L] = E[T] - \frac{Q}{\lambda} = \tau + \frac{1}{\lambda}\left\{E[I(\tau)] - r\right\}. \tag{11}$$

Because the demand rate is λ, the expected lost sales per cycle is given by

$$E[L] = \lambda E(T_L) = \lambda\tau + E[I(\tau)] - r. \tag{12}$$

4. Solution procedure

The optimal control parameters, r, b_1, b_2, t_1 and Q, are found by solving (P1):

$$\text{(P1)} \quad \underset{r,b_1,b_2,t_1,Q}{\text{minimize}} \quad \bar{C}(r, b_1, b_2, t_1, Q),$$
$$\text{subject to} \quad r \geqslant 0, b_2 \geqslant b_1 \geqslant 0,\ 0 \leqslant t_1 \leqslant \tau,\ Q \geqslant r + b_2 + 1,$$
$$r, b_1, b_2 \text{ and } Q \text{ integers},$$

where $\bar{C}(\cdot)$ is given by (1). Following Section 3, the objective function $\bar{C}(\cdot)$ is a cubic equation in Q. Thus, Q^*, the solution of (P1) for given values of r, b_1, b_2 and t_1, can be found analytically. We first assume that Q is a continuous variable and derive the unbounded minimization, Q', of $\bar{C}(\cdot)$, and then we determine the bounded solution, Q^*, considering the constraint $Q \geqslant r + b_2 + 1$, as well as integrality.

We could not show any desirable property of $\bar{C}(\cdot)$ with respect to r, b_1, b_2 or t_1. Thus, $(r^*, b_1^*, b_2^*, t_1^*)$, the optimal solution of the remaining variables, is determined by performing an exhaustive search on r, b_1, b_2 and t_1, while deriving Q^* at each iteration. The proposition below provides a closed form solution for Q^*.

Proposition. *Q^*, For any given value of (r, b_1, b_2, t_1), is an integer that satisfies $Q \geqslant r + b_2 + 1$, either the closest to and below $\tilde{Q}$, or the closest to and above $\tilde{Q}$, where $\tilde{Q} = \max\{Q', r + b_2 + 1\}$, and Q' is the larger of the two roots of the following quadratic equation of Q:*

$$\frac{h}{2\lambda}Q^2 + hE[T_L]Q + \alpha = 0, \tag{13}$$

where

$$\alpha = (\lambda c + hE[I(\tau)])E[T_L] - \frac{h}{2\lambda}\left\{E[I^2(\tau)] - r(r+1)\right\}$$
$$- \{K + \pi E[L] + \hat{\pi}E[B] + \pi' E[B_T] + hE[I_B]\}.$$

With $E[I(\tau)]$ and $E[I^2(\tau)]$ as given by Eq. (A1) *in the appendix, and $E[T_L]$, $E[L]$, $E[B]$, $E[B_T]$ and $E[I_B]$ by Eqs.* (11), (12), (4), (5) *and* (6), *respectively.*

Proof. Let $\tilde{Q}$, be the relaxed solution of (P1) for any given value of (r, b_1, b_2, t_1). To solve this constrained single variable optimization problem we should explore the roots of

$$\partial \bar{C}((r, b_1, b_2, t_1);\ Q)/\partial Q = \frac{\partial E[C]/\partial Q}{E[T]} - \frac{E[C]}{\lambda E[T]^2} = 0, \tag{14}$$

or equivalently, since $E[T] > 0$, the roots of

$$\lambda E[T]\partial E[C]/\partial Q - E[C] = 0. \tag{15}$$

Using

$$\frac{\partial E[C]}{\partial Q} = \frac{h}{\lambda}Q + \left\{c + \frac{hE[I(\tau)]}{\lambda}\right\} \quad \text{and} \quad \frac{\partial E[T]}{\partial Q} = \frac{1}{\lambda},$$

Eq. (15) reduces to the quadratic Eq. (13). We note that the coefficients of both Q and Q^2 are positive in Eq. (13), the derivative of $\bar{C}(\cdot)$ with respect to Q.

Thus, if $\alpha < 0$, then Eq. (13) has exactly one positive real root (Q') and one negative real root. Thus, $\bar{C}(\cdot)$ decreases with Q for $0 \leqslant Q < Q'$ and increases with Q for $Q > Q'$, and hence, $\tilde{Q} = \max\{Q', r + b_2 + 1\}$. Since $r + b_2 + 1$ is a positive integer, only when $\tilde{Q} = Q'$, the closest integer on each side of $\tilde{Q}$, if satisfies $Q \geqslant r + b_2 + 1$, is a candidate for Q^*. Otherwise, $\alpha > 0$, hence

$\bar{C}(\cdot)$ increases with Q for any $Q \geqslant 0$, and thus $\tilde{Q} = r + b_2 + 1$. Here, the roots of (13) are both either negative or complex. Combining both cases we obtain the proposition. □

If $\tilde{Q} < r + b_2 + 1$, then inventory control policies that permit more than one outstanding order should be considered. When $Q' < 0$ or complex, then it is a good idea to question the accuracy of the model parameters in real-life problems.

During the analysis, we conjectured that requiring $b_1 = 0$ does not hurt optimality. The rationale behind it was that for any given policy with $b_1 > 0$ and $t_1 > 0$, there exists another PB2 policy with $b_1 = 0$ and some smaller t_1 and/or larger b_2 such that $E[B]$ and $E[L]$ remain unchanged. Thus, all the remaining cost components remain unchanged, except $E[B_T]$. But $E[B_T]$ must be reduced, since the first, up to b_1, of the backordered demands are now taken later and thus held for a shorter time each, while the remaining ones, up to $b_2 - b_1$, would on the average be taken during the same time period. In our numerical investigations the conjecture was rejected by only a few cases. These cases, however, suggest that in practice assuming $b_1 = 0$ would not involve any significant cost increase.

5. A numerical example

In this section, we examine the behavior of the PB2 policy and its performance versus its special cases. We then examine the sensitivity of this relative performance, to the lost sales, backorder cost parameters and the demand rate. Finally, we observe the shape of the objective function around the optimal solution.

Before we present and discuss the results, two practical comments regarding the solution process are presented. First, to avoid an exhaustive search of r, b_1, b_2 and t_1 we have used a hierarchical Golden-Section search. In our tests it always reached the same solution obtained by an exhaustive search, though we could not identify sufficient conditions for that. Second, we treat t_1 as integer, but the analysis does not really require that. The main advantage of t_1 being discrete is in limiting the required computations. In this case, the basic probability functions can be computed once for any combination of x and λt and then access them from memory as needed throughout the search.

We use the following parameters: $h = 8$, $K = 200$, $\lambda = 2$, $\tau = 10$, $c = 7.5$, $\hat{\pi} = 10$ and $\pi' = 20$. Figs. 3a–d examine the effect of π (from 20 to 300) on the policy of PB1 and PB2. In Fig. 3a the optimal values of the decision variables, r, Q and b of PB1 and r, Q, b_1, b_2, and t_1 of PB2, are presented. The immediate fill rate ($\text{IFR} = 1 - (E[B] + E[L])/\lambda E[T]$) and the total fill rate ($\text{TFR} = 1 - E[L]/\lambda E[T]$) of these solutions are presented in Fig. 3b. The sum $r + b_2 + Q$ for PB2, and $r + b + Q$ for PB1 are presented as well, to demonstrate its relation to the fill rates. In this figure, hollow shapes represent PB2 and solid PB1. Notice that as π (the cost per unit of lost sales) increases, r, b_1, b_2 and Q are used to hedge against lost sales. For $\pi = 40$ and 60, PB1 coincides with PL ($b = 0$), while PB2 employs b_2 for backordering demands, a short time before the replenishment. As a result, we observe its larger gap between TFR and IFR for these values of π. Fig. 3c compares the optimal expected annual cost under PB2 policy, versus the same under PB1, pure lost (PL) and pure backorder (PB) policies. Considering the cost under PB2 as 100%, the marginal savings of PB1 over the best of PL and PB, and of PB2 versus PB1, are demonstrated in

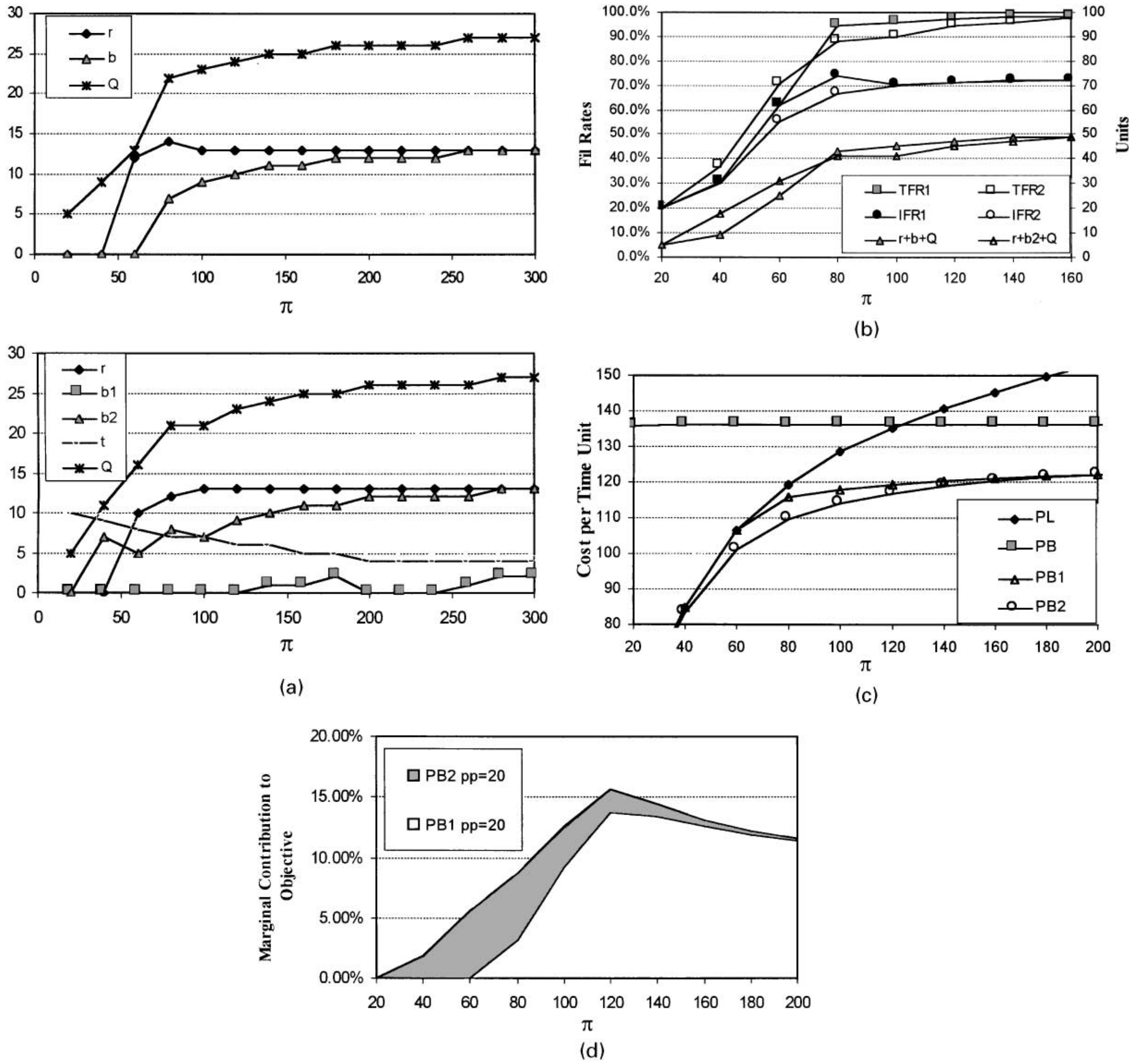


Fig. 3. (a) Optimal solutions of PB1 (upper) and PB2 (lower) versus π (b) Demand fill-rates under PB1 and PB2 (hollow shapes) versus π. (c) Comparison of PB2 with its special cases: PB1, PB and PL, versus π. (d) Contribution of PB2 vs. PB1 and of PB1 vs. PB and PL, as function of π.

Fig. 3d. The main advantage of PB2 is where lost sales cost is too low to justify backordering under PB1 policy.

For $\pi = 120$, the pure lost sales (PL) and the pure backorder (PB) policies switch the lead. At this point the largest cost saving, of more than 15%, is achieved by employing PB2, while employing PB1 can save about 13% (see Figs. 3c and 3d). The largest marginal contribution of PB2 versus PB1 is not at this point, but for $\pi = 60$, where PB1 provides no advantage since it coincides with

PL ($b = 0$ in Fig. 3a). The optimal control variables for PB2 at this point are $r = 10$, $Q = 16$, $b_1 = 0$, $b_2 = 5$, and $t_1 = 8$. The marginal cost saving of PB2 versus PB1 reaches 5.5%, while PB1 provides no advantage over PL. Demonstrating the importance and use of the flexibility provided by the two-segment backorder control limits. For a firm with a large volume of business, 15% and even 5.5% of the inventory related costs could be substantial. In other examples, presented below, the marginal cost saving of PB2 versus PB1 exceeds 7%.

Notice that, our conjecture that $b_1 = 0$ is optimal, is rejected by some cases with $\pi > 120$ (see Fig. 3a). However, the marginal contribution of this variable to the objective is very limited (based on other experiments that are not shown here). Thus, enabling elimination of one dimension in the search with negligible effect on performance. In addition, when $b_1 = 0$ the employment of the policy in practice might be easier.

Another interesting observation is that optimal r under PB2 falls between the optimal r's under PL and PB. We could not prove this result, but it prevailed throughout the numerous experiments we run. This is despite of the fact that as π increases, the values of these r's switch their order. Such a property can be used for limiting the search efforts needed for solving PB2.

As expected, as π increases, the TFR increases (from 20 to 99%) (see Fig. 3b). The increase in IFR is slower, initially IFR $= 20\%$, the same as TFR for $\pi = 20$. The difference TFR–IFR, which demonstrates the percentage of demand backordered, increases up to about 27% at $\pi = 160$, and kept the same at least up to $\pi = 300$ (not shown). When TFR $= 99\%$ at $\pi > 160$, almost no lost sales is tolerated.

For, $\pi < 20$ the optimal PB2 solution is the pure lost-sales policy and for, $\pi > 200$ it slowly reduces to the pure backorder policy. Within the intermediate range, for $20 \leqslant \pi \leqslant 200$, the optimal PB2 policy employs the flexibility provided by the two-segment backorder limits, thus it is not reduced to PB1. The solutions in this range favor lost sales during the first time segment and limited backordering afterwards.

The combined effect of $\hat{\pi}$ (values of 15, 20 and 25) and π (values from 20 to 200) is then examined. The percentage cost savings of PB2 versus PB1, on top of the same of PB1 versus the best of PL and PB, are presented in Fig. 4a ("ph" denotes $\hat{\pi}$). The significant advantage of PB2 is preserved within a wide range of the problem parameters. Similar results can be observed in Fig. 4b for the combined effect of π' (values of 15, 20 and 25) and π (values from 20 to 200) (here "pp" denotes π'). As expected, as the backorder cost parameters increase, the shift from lost sales policy to backorder occurs for larger λ. Fig. 4c demonstrates the effect of λ (values 1, 2 and 3) on the relative advantage of PB2 versus PB1 for various values of π (values from 20 to 200), (in the figure "lam" denotes λ). The sensitivity demonstration shows that the total saving of partial backorder policies might exceed 7% for a wide range of problem parameters.

The shape of the objective function around the optimal solution is presented in Figs. 5a and b. We used the following parameters: $h = 8$, $K = 200$, $\lambda = 2$, $\tau = 10$, $c = 7.5$, $\pi = 80$, $\hat{\pi} = 10$, and $\pi' = 20$. At this point, the optimal control variables for PB2 are $r = 12$, $Q = 21$, $b_1 = 0$, $b_2 = 8$, and $t_1 = 7$. The results demonstrate the difficulty to devise an efficient search procedure. Too large a quantity of r in this example causes a too large inventory, which diminishes the ability to employ the backorder control variables. As a result, we witness a steep increase of the average cost with r. In Fig. 5b we observe a diagonal valley near the optimal solution. As t_1 increases (the second time segment is smaller), decreasing b_2 is preferred. Relatively small deviation in the combination of t_1 and b_2 from optimality, might result in a significant cost increase.

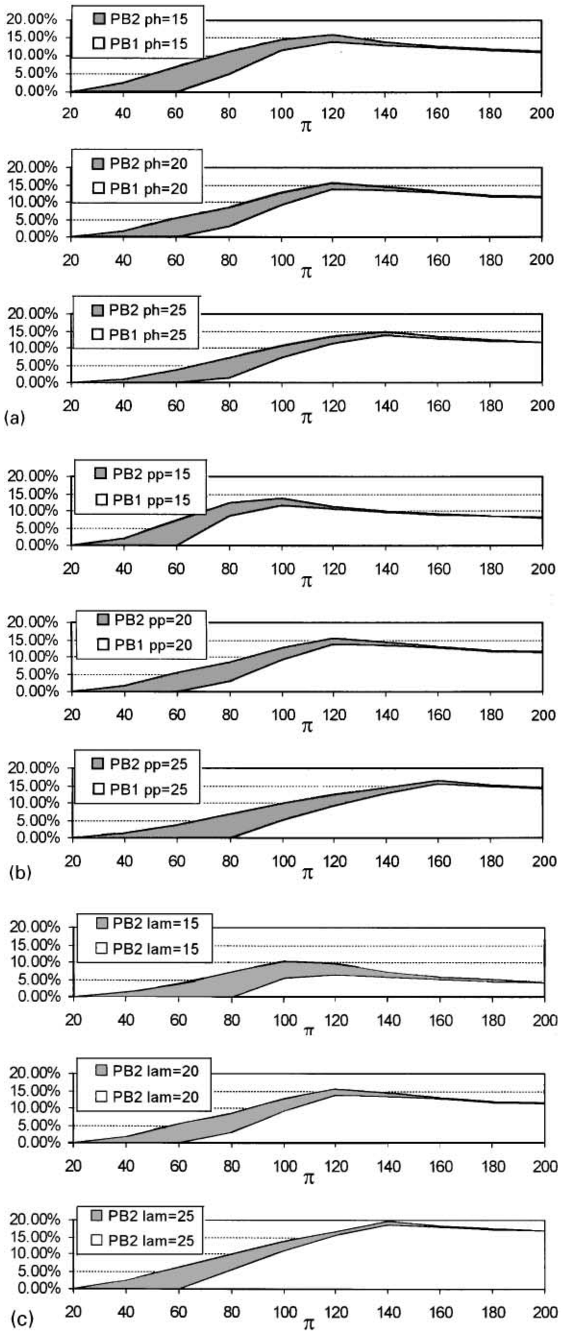
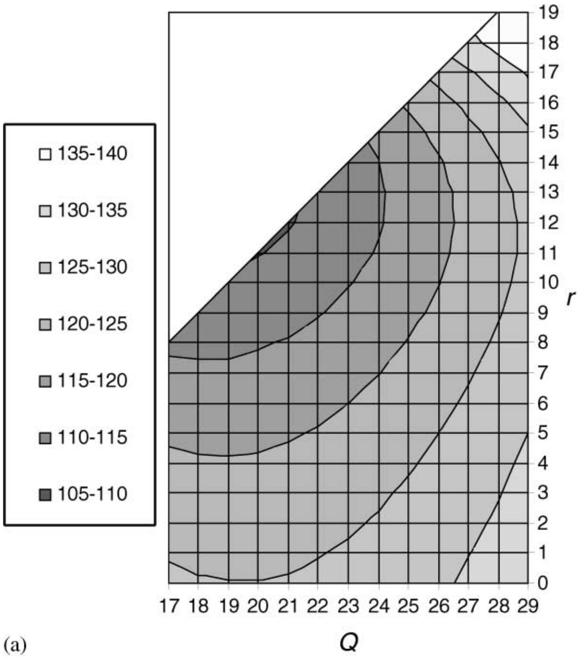
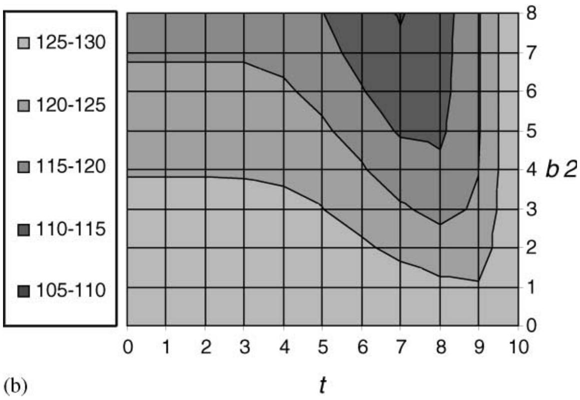


Fig. 4. (a) Contribution of PB2 vs. PB1 and of PB1 vs. PB and PL, as function of $\hat{\pi}$ and π. (b) Contribution of PB2 vs. PB1 and of PB1 vs. PB and PL, as function of π' and π. (c) Contribution of PB2 vs. PB1 and of PB1 vs. PB and PL, as function of λ and π.



(a)



(b)

Fig. 5. (a) The objective function shape around the optimal solution for r and Q. (b) The objective function shape around the optimal solution for t_1 and b_2.

6. Conclusions

It makes sense to assume that the cost of a backordered demand increases when it takes longer to fulfill that demand. Such a time-dependent cost, does not make sense for lost sales. As a result, one might consider losing demands that come early within the lead-time and backorder demands that appear closer to the replenishment. This is the main idea underlying the two-segment partial backorder policy (PB2). Some large chemical corporations employ this inventory management policy and we felt it worth analysis. Due to proprietary restrictions we cannot disclose further corporate information.

The traditional backorder policy, the lost sales policy, and the one-segment partial backorder policy (PB1) are all special cases of PB2, which provides even additional flexibility. Hence, a PB2 policy provides a unified framework for studying policies that combine shortage and backorder.

In this paper we have defined, formulated, solved and demonstrated PB2 and numerically explored the behavior of this policy with respect to various cost parameters. A closed form solution is provided for the order quantity, in regards to any given values of the remaining control variables. We also conjectured, that the optimal backorder limit during the first time segment should be zero. The conjecture was rejected but we found out that setting $b_1 = 0$, would not significantly harm the objective.

A numerical example demonstrates that the PB2 policy can achieve a cost saving of more than 7% over the best of pure backorder (PB), pure lost sales (PL) and PB1 policies. The cost saving of PB2 over the best of PB and PL might exceed 15%. Another important finding is that the largest cost saving occurs at the *breakeven point* where the pure backorder policy and the pure lost sales policy yield identical expected annual costs. The advantage of PB2 is significant over a wide range of the problem parameters. The merits of PB2 is for small values of lost sales cost (π). There, while PB1 coincides with the PL policy, PB2 provides economically justified opportunity for backordering some of the otherwise, lost sales.

Future works will incorporate solution strategy for finding optimal continuous-valued t_1 together with the integer-valued b_1, b_2, r, and Q. An efficient computational procedure still needs to be developed. This includes establishing bounds for b_1 and b_2 to facilitate the search for optimal solutions, and using a more efficient search technique. In addition, general demand processes other than Poisson such as compound Poisson, compound renewal or Markov renewal demand processes may be used. Also, partial backorder policies that utilize more than two time segments of backorder control limits could be explored. Finally, we intend to further develop and test applications of the model with "real" data from companies that face this type of problem. With the chemical company we work with, an inventory cost saving of 2–4% is expected, by implementing the model in its inventory control system.

Acknowledgements

The detailed and constructive comments from the referees have improved the paper and are gratefully acknowledged.

Appendix

Expected inventory level at the replenishment time: The state space $I(\tau)$, is $\{-b_2, -b_2+1, \ldots, -b_1, \ldots, 0, \ldots, r-1, r\}$. Its kth moment can be derived as follows:

$$\begin{aligned} E[I(\tau)^k] &= E_{D_1}[E(I(\tau)^k|D_1)] \\ &= Pr[D_1 \leqslant r+b_1-1]E_{D_1}[E(I(\tau)^k|D_1 \leqslant r+b_1-1)] \end{aligned}$$

$$+ Pr[D_1 \geqslant r + b_1]E_{D_1}[E(I(\tau)^k|D_1 \geqslant r + b_1)]$$

$$= \sum_{i=0}^{r+b_1-1} p(i;\ \lambda t_1)\left\{\sum_{j=0}^{b_2-1+r-i} (r-i-j)^k p(j;\ \lambda(\tau - t_1))\right.$$

$$\left. + (-b_2)^k P(b_2 + r - i;\lambda\ (\tau - t_1))\right\}$$

$$+ P(r + b_1;\ \lambda t_1)\left\{\sum_{j=0}^{b_2-b_1-1} (-b_1 - j)^k p(j;\ \lambda(\tau - t_1))\right.$$

$$\left. + (-b_2)^k P(b_2 - b_1;\ \lambda(\tau - t_1))\right\}. \tag{A.1}$$

Expected number of backorders: The number, B, of backorders in a cycle, is given by $E[B] = EB_1] + E_{D_1}[E(B_2|D_1)]$, where B_1 is the random number of backorders during $(0, t_1]$, and B_2 during $(t_1,\tau]$.

$$E(B_1) = \sum_{i=r+1}^{r+b_1-1} (i-r)p(i;\ \lambda t_1) + b_1 P(r + b_1;\ \lambda t_1) \tag{A.2}$$

$$E[B_2] = E_{D_1}[E(B_2|D_1)] = Pr[D_1 \leqslant r]E_{D_1}[E(B_2|D_1 \leqslant r)]$$

$$+\ Pr[r + 1 \leqslant D_1 \leqslant r + b_1 - 1]E_{D_1}[E(B_2|r + 1 \leqslant D_1 \leqslant r + b_1 - 1)]$$

$$+\ [D_1 \geqslant r + b_1 > r]E_{D_1}[E(B_2|D_1 \geqslant r + b_1 > r)]$$

$$= \sum_{i=0}^{r} p(i;\ \lambda t_1)\left\{\sum_{j=1}^{b_2-1} jp(j + r - i;\ \lambda(\tau - t_1)) + b_2 P(b_2 + r - i;\ \lambda(\tau - t_1))\right\}$$

$$+ \sum_{i=r+1}^{r+b_1-1} p(i;\ \lambda t_1)\left\{\sum_{j=1}^{b_2-1+r-i} jp(j;\ \lambda(\tau - t_1))\right.$$

$$\left. + (b_2 + r - i)P(b_2 + r - i;\ \lambda(\tau - t_1))\right\}$$

$$+\ P(r + \delta;\ \lambda t_1)\left\{\sum_{j=1}^{b_2-b_1-1} jp(j;\ \lambda(\tau - t_1))\right.$$

$$\left. + (b_2 - b_1)P(b_2 - b_1;\ \lambda(\tau - t_1))\right\}. \tag{A.3}$$

where $\delta = \max(b_1, 1)$.

Expected time-dependent backorder per cycle: The expected time-dependent backorder, $E[B_T]$, is defined by Eq. (5). Let us first define the following two expressions:

$$g(x,\lambda,\omega) = \int_0^{\omega} p(x,\ \lambda t)\,\mathrm{d}t = \frac{1}{\lambda}\left(1 - \mathrm{e}^{-\lambda\omega}\sum_{y=0}^{x}\frac{(\lambda\omega)^{(x-y)}}{(x-y)!}\right) = \frac{P(x+1,\lambda\omega)}{\lambda}$$

and

$$G(x,\lambda,\omega)=\int_0^{\omega}P(x,\ \lambda t)\,\mathrm{d}t=\omega-\sum_{y=0}^{x-1}g(y,\ \lambda,\ \omega)=\sum_{y=0}^{x}P(y,\ \lambda\omega)/\lambda. \tag{A.4}$$

The first component of (5) is given by

$$\begin{aligned}\int_0^{t_1}E[B_1(t)]\,\mathrm{d}t&=\int_0^{t_1}\left\{\sum_{i=r+1}^{r+b_1-1}(i-r)p(i;\ \lambda t)+b_1P(b_1+r;\ \lambda_t\right\}\mathrm{d}t\\&=\sum_{i=r+1}^{r+b_1-1}(i-r)\int_0^{t_1}p(i;\ \lambda t)\,\mathrm{d}t+b_1\int_0^{t_1}P(b_1+r;\ \lambda t)\,\mathrm{d}t\\&=\sum_{i=r+1}^{r+b_1-1}(i-r)g(i,\ \lambda,\ t_1)+b_1G(b_1+r,\ \lambda,\ t_1).\end{aligned} \tag{A.5}$$

The second component of (5) can be decomposed into three parts as follows:

$$\begin{aligned}\int_{t_1}^{\tau}E[B_2(t)]\,\mathrm{d}t&=E_{D_1}\left[\int_{t_1}^{\tau}E(B_2(t)|D_1)\,\mathrm{d}t\right]\\&=Pr[D_1\leqslant r]E_{D_1}\left[\int_{t_1}^{\tau}E(B_2(t)|D_1\leqslant r)\,\mathrm{d}t\right]\\&\quad+\ Pr[r+1\leqslant D_1\leqslant r+b_1-1]E_{D_1}\\&\quad\times\left[\int_{t_1}^{\tau}E(B_2(t)|r+1\leqslant D_1\leqslant r+b_1-1)\,\mathrm{d}t\right]\\&\quad+\ Pr[D_1\geqslant r+b_1>r]E_{D_1}\left[\int_{t_1}^{\tau}E(B_2(t)|D_1\geqslant r+b_1>r)\,\mathrm{d}t\right]\\&=\sum_{i=0}^{r}p(i;\ \lambda t_1)\left\{\sum_{j=1}^{b_2-1}j\int_0^{\tau-t_1}p(\ j+r-i;\ \lambda t)\,\mathrm{d}t+b_2\int_0^{\tau-t_1}P(b_2+r-i;\ \lambda t)\,\mathrm{d}t\right\}\\&\quad+\sum_{i=r+1}^{r+b_1-1}p(i;\ \lambda t_1)\left\{(i-r)(\tau-t_1)+\sum_{j=1}^{b_2-1+r-i}j\int_0^{\tau-t_1}p(j;\ \lambda t)\,\mathrm{d}t\right.\\&\quad\left.+\,(b_2+r-i)\int_0^{\tau-t_1}P(b_2+r-i;\ \lambda t)\,\mathrm{d}t\right\}\\&\quad+\ P(r+\delta;\ \lambda t_1)\left\{b_1(\tau-t_1)+\sum_{j=1}^{b_2-b_1-1}j\int_0^{\tau-t_1}p(j;\ \lambda t)\,\mathrm{d}t\right.\\&\quad\left.+\,(b_2-b_1)\int_0^{\tau-t_1}P(b_2-b_1;\ \lambda t)\,\mathrm{d}t\right\}\\&=\sum_{i=0}^{r}p(i;\ \lambda t_1)\left\{\sum_{j=1}^{b_2-1}jg(j+r-i,\ \lambda,\ \tau-t_1)+b_2G(b_2+r-i,\lambda,\tau-t_1)\right\}\end{aligned}$$

$$+\sum_{i=r+1}^{r+b_1-1} p(i;\,\lambda t_1)\Big\{(i-r)(\tau-t_1)+\sum_{j=1}^{b_2-1+r-i} jg(j,\,\lambda,\,\tau-t_1)$$
$$+(b_2+r-i)G(b_2+r-i,\lambda,\,\tau-t_1)\Big\}+P(r+\delta;\,\lambda t_1)$$
$$\times\Big\{b_1(\tau-t_1)+\sum_{j=1}^{b_2-b_1-1} jg(j,\,\lambda,\,\tau-t_1)$$
$$+(b_2-b_1)G(b_2-b_1,\lambda,\tau-t_1)\Big\} \quad \text{(A.6)}$$

Expected inventory time after replenishment: Let us derive $E[I_A]$ as given in Eq. (7). The quantity $Q+I(\tau)$ is the random inventory level just after replenishment. The time between arrivals of demand events is exponentially distributed with mean $1/\lambda$. Thus for a given $I(\tau)$, the expected accumulated inventory time from a replenishment epoch until the next ordering time is given by

$$E[I_A|I(\tau)=x]=\frac{1}{\lambda}\sum_{y=0}^{x+Q-r-1}(x+Q-y)=\frac{1}{2\lambda}\Big[(x+Q)^2-r(r+1)\Big]. \quad \text{(A.7)}$$

Hence,

$$E[I_A]=E_{I(\tau)}[E[I_A|I(\tau)]]=\sum_x Pr(I(\tau)=x)E(I_A|I(\tau)=x]$$
$$=\frac{1}{2\lambda}\Big\{E[I(\tau)^2]+2QE[I(\tau)]+Q^2-r(r+1)\Big\}, \quad \text{(A.8)}$$

where the moments of $I(\tau)$ are given by (A.1).

References

[1] Peterson R, Silver EA. Decision systems for inventory management and production planning, 2nd ed.. New York: Wiley, 1995.
[2] Rabinowitz G, Mehrez A, Chu CW, Patuwo BE. A partial backorder control for continuous review (r, Q) inventory system with poisson demand and constant lead time. Computers and Operations Research 1995;22(7):689–700.
[3] Montgomery DC, Baser MS, Baser X, Caisson AK. Inventory models with a mixture of backorders and lost sales. Naval Research Logistics Quarterly 1981;20:255–63.
[4] Rosenberg D. A new analysis of lot-size model with partial backlogging. Naval Research Logistics Quarterly 1979;26:349–53.
[5] Kim DH, Park KS. (Q,R) inventory model with a mixture of lost sales and time weighted backorders. J Operational Research Society 1985;36:231–8.
[6] Woo YY, Sphicas GP. (Q, R) inventory models with partial backorders, poisson demands, and exponential lead times. Dept. of Management Science, National Chico Tong University, Hsinchu, Taiwan, 1991.
[7] Posner MJM. Class of inventory models with customer impatience. Naval Research Logistics Quarterly 1972;19:483–92.
[8] Das C. The (S − 1, S) inventory model under time limit on backorders. Operations Research 1977;25:835–50.
[9] Moinzadeh K. Operating characteristics for the (S − 1, S) inventory system with partial backorders and constant supply time. Management Science 1989;35:472–7.
[10] Smeitink E. A note on the operating characteristics of the (S − 1, S) inventory systems with partial backorders and constant resupply times. Management Science 1990;36:1413–4.

[11] Porteus EL. Stochastic Inventory Theory. In: Heyman X, Sobel X, editors. Handbook in operations research and management science, vol. 2. Amsterdam: North-Holland, 1990.
[12] Nahmias S. Simple approximation for a variety of dynamic leadtime lost-sales inventory models. Operations Research 1979;27:904–24.
[13] Zheng YS, Federgruen A. Finding optimal (s, S) policy is about as simple as evaluating a single policy. Operations Research 1991;39:654–65.
[14] Zheng YS, Federgruen A. Corrections to: finding optimal (s, S) policy is about as simple as evaluating a single policy. Operations Research 1992;40:192.
[15] Federgruen A, Zheng YS. An efficient algorithm for computing an optimal (r, Q) policy in continuous review stochastic inventory systems. Operations Research 1992;40:808–13.
[16] Zheng YS. On properties of stochastic inventory systems. Management Science 1992;38:87–103.
[17] Cinlar E. Introduction to stochastic processes. Englewood Cliffs, NJ: Prentice-Hall Inc, 1975.
[18] Hadley G, Whitin TM. Analysis of Inventory Systems. Englewood Cliffs, NJ: Prentice-Hall, 1963.
[19] Ross S. Stochastic Processes. New York: Wiley, 1983.

Ching-Wu Chu earned his Ph.D. from the College of Business at Kent State University. His research interests focus on the theory and practice of operation and scheduling of production and inventory systems.

B. Eddy Patuwo is an Associate Professor in the Dept. of Administrative Sciences. He had a Ph.D. in IEOR from Virginia Polytechnic Institute and State University. His research interests are in the study of neural networks and stochastic models for production/inventory systems. He has published in *IIE Transactions*, *Decision Sciences*, *Journal of the Operational Research Society*, *European Journal of Operations Research*, *Computers & Operations Research*, *International Jouranl of Forecasting*, and others.

Abraham Mehrez serves as the head of the Industrial Engineering and Management Department at Ben-Gurion University. He earned his Ph.D. from Johns Hopkins University. His research is in interdisciplinary management research and applications on the analysis of public and private sector systems. He has about 200 publications in these areas and he is in the editorial advisory board of *Computers and Operations Research*, and area editor in *Production and Operations Management* and an associate editor in *Management Science*.

Gad Rabinowitz is on the faculty of the Industrial Engineering and Management Department at Ben Gurion University. He earned his Ph.D. in Operations Research from Case Western Reserve University. His research interests focus on the theory and practice of operation and scheduling of production and logistics systems; modeling of quality engineering and management issues; quality management in health care systems; and the design and operation of multi-quality water supply systems.