Journal of the Operational Research Society (1998) 49, 509–518
http://www.stockton-press.co.uk/jor

Rationing policies for some inventory systems

I Moon[1] and S Kang[2]

[1]*Pusan National University and* [2]*Seoul National University, Korea*

This paper considers inventory systems which maintain stocks to meet various demand classes with different priorities. We use the concept of a support level control policy. That is rationing is accomplished by maintaining a support level, say K, such that when on hand stock reaches K, all low priority demands are backordered. We develop four analytical and simulation models to improve the existing models. Firstly, multiple support levels are used instead of using a single support level. Secondly, a simulation model with a more realistic assumption on the demand process has been provided. Thirdly, a single period deterministic cost minimisation model has been developed analytically. Finally, we address a continuous review (Q, r) model with a compound Poisson process.

Keywords: inventory; rationing policy; simulation

Introduction

Distinguishing between various priority classes of demand arises frequently in practice. For example in a hospital emergency room blood is rationed in accordance to the emergency levels of patients. A second example is a computer monitoring system where computer codes and blocks are allocated in accordance with the priority levels such as professors, graduate students and undergraduate students, etc. A third example is a military system where an item is used in several different weapons or by units with different missions. A fourth example is a general sales company where different customers for the same product yield different profits per unit sold. In addition to these systems, we find examples of rationing in airline reservations, pre-sales of season tickets, etc.

Several researchers have dealt with inventory systems in which there are several priority customers. Kaplan[1] addresses the use of reserve levels, that is stock levels at which to stop issuing in response to lower priority customer. Veinott[2] develops a multiperiod single product nonstationary inventory model where the system is reviewed at the beginning of each of a sequence of periods of equal length. Evans[3] and Topkis[4] independently develop conditions where optimal rationing policy between successive procurements of new stock is determined by a set of critical rationing levels such that at a given time one satisfies demand of a given class only if no demand of a more important class remains unsatisfied and as long as the stock level does not fall below the critical rationing level for that class at that time. Nahmias and Demmy[5] develop methods to compute the expected number of backorders for high and low priority customers for several inventory models. They develop methods to show how the support level, that is, the inventory level that starts rationing for low priority customers, reorder point, and order quantity behave to meet a desired fill rate. However, they do not consider a cost optimisation model. Haynsworth and Price[6] study a periodic review model with a discrete time rationing policy with a desired service level for high-priority demands. They model the way reserve levels vary with remaining lead time and show that the system is more effective than that of using a single reserve level to be used throughout lead time. Recently Ha[7] studies a rationing problem of a make-to-stock production system with several demand classes. A queueing model has been developed to compare the optimal rationing policy with the FCFS policy. Note that the reserve level, the rationing level, and the control level are equivalent terms used in various literature. The fill rate also has the same meaning with the service level.

The goal of this paper is to improve and extend existing studies, especially the models of Nahmias and Demmy[5]. The concept of support level control policy is used in this paper. Rationing is accomplished by maintaining *a support level*, say K, such that when on hand stock reaches K, all low priority demands are backordered. Firstly, we extend the single period model with two demand classes (that is a single support level) developed by Nahmias and Demmy[5] to that of multiple demand classes (that is multiple support levels) in the next section. However, the model developed is based on the assumption that all demands occur simultaneously at the end of the planning period. It is more reasonable to use the assumption that demand occurs uniformly during the period even though it is impossible to obtain analytical solutions for this case. In the subsequent section, we develop a simulation model under this better assumption. No cost consideration is given in the two sections. There then follows a section developing a single period cost optimisation model under the assumption that

Correspondence: Dr I Moon, Department of Industrial Engineering, Pusan National University, Pusan 609-735, Korea.
E-mail: ikmoon@hyowon.pusan.ac.kr

the demand rate from each class is constant. However, we do not assume, as in the second section, that all demands occur at the end of the planning horizon. Instead, holding cost is charged per unit per unit time. We analyse the tradeoff between holding and shortage cost for each demand class. We show that the total cost during the planning horizon is convex in *rationing trigger times*, that is the times at which the system starts to reject customers. We develop a simple sequential procedure to find the optimal rationing trigger times. We show that small changes in the available initial inventory result in equal shifts of the optimal rationing trigger times. The rationing policies obtained in the section might be coupled with some safety stock strategies for the random demand case. Finally, we extend the continuous review model with Poisson demands developed by Nahmias and Demmy[5] to that of compound Poisson demands. Customer fill rates are used to compare the effects of rationing for the model. The paper concludes with a brief summary of its main points and possible research areas.

A single period model with multiple support levels

In this section, we extend the single period model developed by Nahmias and Demmy[5] to one with multiple support levels. Suppose at the start of a single planning period, the level of starting stock is J and the support levels are $K_1, K_2, \ldots, K_{N-1}$ where $0 \leqslant K_1 \leqslant K_2 \leqslant \cdots \leqslant K_{N-1} \leqslant J$. Furthermore suppose that the total periodic demand, D, can be decomposed into $D = D_1 + D_2 + \cdots + D_N$ where D_i is demand whose priority is i (Let priority 1 be the highest priority). Rationing is accomplished by maintaining support levels, K_is, such that when on hand stock reaches K_i all demand classes whose priorities are lower than i are backordered. Assume that D_i has a known cumulative distribution function F_i and a density function f_i for all i, and further that D_is occur simultaneously at the end of each planning period. As Nahmias and Demmy[5] pointed out, one can consider this assumption as an approximation to the more reasonable assumption that demand occurs uniformly during the planning period.

Define Z_i as the number of backorders for demand class i at the end of the period. Then

$$Z_1 = (D_1 - J)^+ \tag{1}$$

and for $i \geqslant 2$

$$Z_i = \begin{cases} D_i \text{ if } J - (D_1 + \cdots + D_{i-1}) < K_{i-1} \\ (D_1 + \cdots + D_i - J + K_{i-1})^+ \text{ otherwise} \end{cases} \tag{2}$$

where $x^+ \equiv \max(x, 0)$. Then the expected number of backorders for each demand class can be represented as follows:

$$B_1 = \int_J^\infty (D_1 - J) f_1(D_1) dD_1 \tag{3}$$

and for $i \geqslant 2$

$$\begin{aligned} B_i = E(Z_i) &= E(D_i) P[J - (D_1 + \cdots + D_{i-1}) < K_{i-1}] \\ &+ E[(D_1 + \cdots + D_i - J + K_{i-1})^+, \\ &\qquad J - (D_1 + \cdots + D_{i-1}) \geqslant K_{i-1}] \end{aligned} \tag{4}$$

where

$$\begin{aligned} &E[(D_1 + \cdots + D_i - J + K_{i-1})^+, \\ &\qquad J - (D_1 + \cdots + D_{i-1}) \geqslant K_{i-1}] \\ &= \int_{J-D_1 \geqslant K_{i-1}, \cdots, J-(D_1+\cdots+D_{i-1}) \geqslant K_{i-1}, D_1+\cdots+D_i \geqslant J-K_{i-1}} \cdots \\ &\int (D_1 + \cdots + D_i - J + K_{i-1}) f_1(D_1) \cdots f_i(D_i) dD_1 \cdots dD_i \\ &= \int_0^{J-K_{i-1}} f_1(D_1) \cdots \int_0^{J-(D_1+\cdots+D_{i-2})-K_{i-1}} f_{i-1}(D_{i-1}) \\ &\int_{J-K_{i-1}-(D_1+\cdots+D_{i-1})}^{J-K_{i-1}} (D_1 + \cdots + D_i - J + K_{i-1}) \\ &\quad \times f_i(D_i) dD_i \cdots dD_1 \end{aligned} \tag{5}$$

Example 1 There is an inventory system which classifies demand as three classes. Suppose that three priority demands are given by independent exponential random variables with parameter λ. The above equations give:

$$B_1 = \frac{1}{\lambda} e^{-\lambda J}$$

$$B_2 = \left(J - K_1 + \frac{1}{\lambda}\right) e^{-\lambda(J-K_1)}$$

$$B_3 = \left[J - K_2 + \frac{1}{\lambda} + \frac{\lambda(J-K_2)^2}{2}\right] e^{-\lambda(J-K_2)}$$

When there is no rationing, total demand, D, will have the Erlang-3 distribution with parameter λ. Consequently, the expected backorders without rationing is as follows:

$$B = E[(D-J)^+] = \left(2J + \frac{3}{\lambda} + \frac{\lambda J^2}{2}\right) e^{-\lambda J}$$

We present computational results for rationing and no rationing for $\lambda = 0.1$ and $J = 50$ in Table 1. Notice that B_1 is independent of support levels since the model assumes that all demands occur simultaneously at the end of the

Table 1 Expected number of backorders with rationing and without rationing

K_2	K_1	B_1	B_2	B_3	B
40	0	0.0674	0.4043	9.1970	1.7182
40	10	0.0674	0.9158	9.1970	1.7182
40	20	0.0674	1.9915	9.1970	1.7182
40	30	0.0674	4.0601	9.1970	1.7182
30	0	0.0674	0.4043	6.7668	1.7182
30	10	0.0674	0.9158	6.7668	1.7182
30	20	0.0674	1.9915	6.7668	1.7182
20	0	0.0674	0.4043	4.2319	1.7182
20	10	0.0674	0.9158	4.2319	1.7182
10	0	0.0674	0.4043	2.3910	1.7182
0	0	0.0674	0.4043	1.2465	1.7182

planning period. One of the disadvantages of this model is that we can not see the effect of changing control levels. We will develop a simulation model which overcomes this problem in the next section.

A single period model with Poisson demand

As stated in the previous section, Nahmias and Demmy[5] developed an inventory model with a single support level with the assumption that all demands occur simultaneously at the end of the planning period. As a result of using this unrealistic assumption, the expected backorders B_1 and B_2 have no correlation as a support level, K, changes (See Table 1). As pointed out by Nahmias and Demmy[5], it is more reasonable to use the assumption that demand occurs uniformly during the period. It means that the arrival process follows a Poisson process. However, we cannot obtain analytical solutions for this case, and leave this as an open research problem. Consequently, we develop a simulation model under this assumption. The flowchart for the simulation algorithm has been depicted in Appendix A, and the pseudo-code of the algorithm has been presented in Appendix B. The model has been coded using Microsoft Visual Basic 5.0, and the simulation experiments have been run for 100 replications. The computational results obtained from the simulation model are represented in Table 2 to be compared with those of Nahmias and Demmy[5]. Note that both types of expected backorders are affected by the change of a support level. Since both types of expected backorders are affected by changing a support level, the results show the trade-off between high and low priority demands and are therefore more relevant to decision makers.

A single period model with rationing trigger times

In this section, we assume the demand rate from each class is constant. However, we no longer assume that all demands occur at the end of the planning horizon, T. Instead, holding cost, h, is charged per unit per unit time. We use *rationing trigger times* rather than support levels as decision variables. This makes the mathematical derivations easier. Also, customers are interested in the *time epoch*, and not in the support level, that triggers rationing. The rationing policies obtained in this section might be coupled with safety stocks for the random demand case.

Let ρ_i be the cost per unit lost sales to customer type i. Without loss of generality, we reorder customer types so that $\rho_1 \leqslant \rho_2 \leqslant \cdots \leqslant \rho_N$. The objective is to minimise the holding and penalty costs. This model is also equivalent to one in which demand classes are distinguished by the prices at which the items are sold, and is recognized, for example, as a variant of the airline overbooking problem. In this case, the objective is to maximise net profit which equals revenue minus holding cost. We use a rationing policy such that demands of customer type i are satisfied up to time $\tau_i (\in [0, T])$, $i = 1, \ldots, N$. We assume that initial inventory $J < T\sum_{i=1}^{N} D_i$. If $J \geqslant T\sum_{i=1}^{N} D_i$, then clearly $\tau_i^* = T$ for all i, that is we do not ration at all. We can prove easily that optimal rationing trigger times τ_is are nondecreasing in i. That is, there exists no optimal rationing trigger times such that $\tau_i > \tau_j$ where $i < j$. Since $\sum_{j=1}^{N} D_j\tau_j = J$, we can eliminate τ_N by noting that $\tau_N = (J - \sum_{j=1}^{N-1} D_j\tau_j)/D_N$. We seek to find optimal rationing trigger times $\tau_1^*, \ldots, \tau_{N-1}^*$ and hence τ_N^* which minimises total cost.

Table 2 Computational results for the simulation study

			Nahmias and Demmy		*Simulation*	
J	K	B	B_1	B_2	B_1	B_2
50	0	0.4717	0.0674	0.4043	0.3181	0.3154
50	10	0.4717	0.0674	0.9158	0.1895	0.7431
50	20	0.4717	0.0674	1.9915	0.1410	1.6594
50	30	0.4717	0.0674	4.0601	0.1147	3.5348
50	40	0.4717	0.0674	7.3576	0.0988	6.8902
100	0	0.0054	0.0005	0.0050	0.0064	0.0056
100	10	0.0054	0.0005	0.0123	0.0046	0.0114
100	20	0.0054	0.0005	0.0302	0.0031	0.0218
100	30	0.0054	0.0005	0.0730	0.0021	0.0522
100	40	0.0054	0.0005	0.1735	0.0015	0.1265
100	50	0.0054	0.0005	0.4043	0.0008	0.3154
100	60	0.0054	0.0005	0.9158	0.0005	0.7431
100	70	0.0054	0.0005	1.9915	0.0002	1.6594
100	80	0.0054	0.0005	4.0601	0.0000	3.5348
100	90	0.0054	0.0005	7.3576	0.0000	6.8902

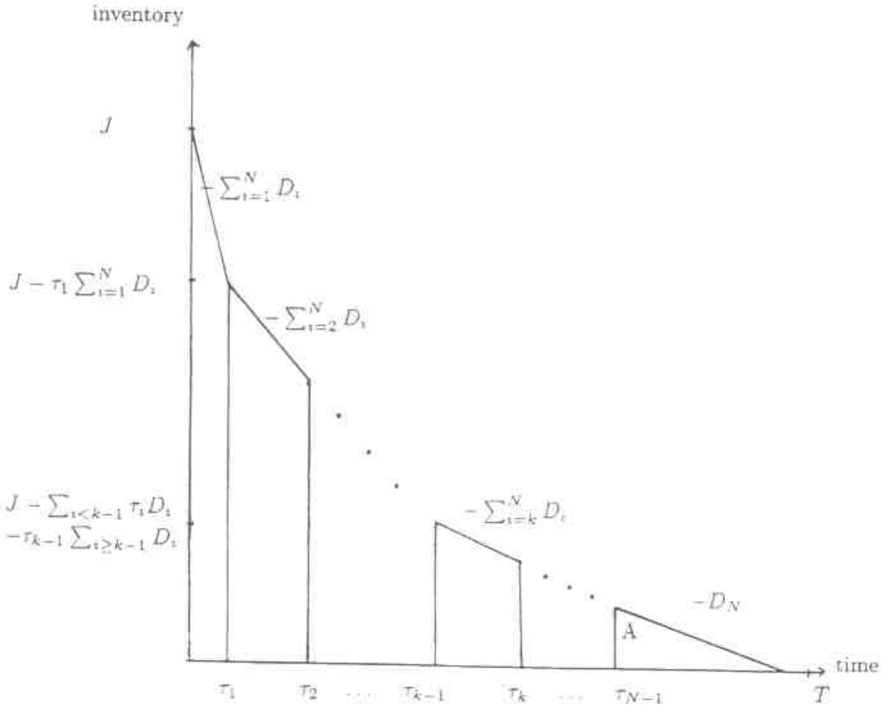


Figure 1 The relationship between inventory and rationing trigger times.

We compute the holding cost as follows: (See Figure 1) Firstly, the area of trapezoids is

$$\sum_{k=1}^{N-1} \frac{1}{2}(\tau_k - \tau_{k-1}) \left\{ \left[J - \sum_{i<k-1} \tau_i D_i - \tau_{k-1} \sum_{i \geqslant k-1} D_i \right] + \left[J - \sum_{i<k} \tau_i D_i - \tau_k \sum_{i \geqslant k} D_i \right] \right\}$$
$$= \frac{1}{2} \left[\sum_{i=1}^{N-2} \tau_i^2 D_i - \tau_{N-1}^2 (D_{N-1} + D_N) + 2\tau_{N-1} J - 2\tau_{N-1} \left(\sum_{i=1}^{N-2} \tau_i D_i \right) \right] \quad (6)$$

The area of triangle A is

$$\frac{1}{2D_N} \left(J - \sum_{i<N-1} \tau_i D_i - \tau_{N-1} \sum_{i \geqslant N-1} D_i \right)^2 \quad (7)$$

Consequently, the holding cost during the horizon becomes

$$\frac{h}{2} \left\{ \sum_{i=1}^{N-2} \tau_i^2 D_i - \tau_{N-1}^2 (D_{N-1} + D_N) + 2\tau_{N-1} J - 2\tau_{N-1} \left(\sum_{i=1}^{N-2} \tau_i D_i \right) + \frac{1}{D_N} \left[J - \sum_{i<N-1} \tau_i D_i - \tau_{N-1}(D_{N-1} + D_N) \right]^2 \right\} \quad (8)$$

The penalty cost during the horizon is

$$\sum_{i=1}^{N-1} \rho_i D_i (T - \tau_i) + \rho_N D_N \times \left[T - \frac{J - \sum_{i<N-1} \tau_i D_i - \tau_{N-1}(D_{N-1} + D_N)}{D_N} - \tau_{N-1} \right] \quad (9)$$

The total holding and penalty cost during the horizon becomes

$$C(\tau_1, \ldots, \tau_{N-1}) = \sum_{i=1}^{N-1} \frac{hD_i}{2} \left(1 + \frac{D_i}{D_N} \right) \tau_i^2 + \sum_{i<j} \frac{hD_i D_j}{D_N} \tau_i \tau_j + \sum_{i=1}^{N-1} \left(\rho_N - \rho_i - \frac{hJ}{D_N} \right) D_i \tau_i + \frac{hJ^2}{2D_N} + T \sum_{i=1}^{N} \rho_i D_i - \rho_N J \quad (10)$$

Noting that $C(\tau_1, \ldots, \tau_{N-1})$ is quadratic, we rewrite $C(\tau_1, \ldots, \tau_{N-1})$ as $\frac{1}{2} \underline{\tau}^{\mathrm{T}} H \underline{\tau} - \underline{g}^{\mathrm{T}} \underline{\tau}$ + constant where

$$H = h \operatorname{diag}(D_i) + \frac{h}{D_N} \underline{D}\, \underline{D}^{\mathrm{T}}, \quad (11)$$

and

$$\underline{g} = \left(\frac{hJ}{D_N} - \rho_N \right) \underline{D} + \operatorname{diag}(\rho) \underline{D}, \quad (12)$$

where $\operatorname{diag}(D_i)$ denotes an $(N-1) \times (N-1)$ diagonal matrix, $\underline{D} = (D_1, \ldots, D_{N-1})^{\mathrm{T}}$ and $\underline{\tau} = (\tau_1, \ldots, \tau_{N-1})^{\mathrm{T}}$.

First we prove that $C(\underline{\tau}^{\mathrm{T}})$ is convex in rationing trigger times.

Property 1 *$C(\underline{\tau}^{\mathrm{T}})$ is convex in $\tau_1, \ldots, \tau_{N-1}$.*

Proof Equivalently, we show that the following Hessian matrix

$$\nabla^2 C(\underline{\tau}^{\mathrm{T}}) = \begin{pmatrix} hD_1\left(1+\frac{D_1}{D_N}\right) & hD_1\frac{D_2}{D_N} & hD_1\frac{D_3}{D_N} & \cdots & hD_1\frac{D_{N-1}}{D_N} \\ hD_2\frac{D_1}{D_N} & hD_2\left(1+\frac{D_2}{D_N}\right) & hD_2\frac{D_3}{D_N} & \cdots & hD_2\frac{D_{N-1}}{D_N} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ hD_{N-1}\frac{D_1}{D_N} & hD_{N-1}\frac{D_2}{D_N} & hD_{N-1}\frac{D_3}{D_N} & \cdots & hD_{N-1}\left(1+\frac{D_{N-1}}{D_N}\right) \end{pmatrix}$$

is positive definite. This follows immediately from the fact that n_{th} $(n = 1, \ldots, N-1)$ principal minor is $h^n D_1 D_2 \ldots D_n \sum_{i=1}^{n+1} D_i / D_{n+1} > 0$. □

The unrestricted minimum of $C(\underline{\tau}^{\mathrm{T}})$ is given by $\underline{\tau}^0 = H^{-1}\underline{g}$. Using the Sherman–Morrison Matrix Identity, we obtain

$$H^{-1} = \frac{1}{h}\left[\mathrm{diag}\left(\frac{1}{D_i}\right) - \frac{\underline{e}\,\underline{e}^{\mathrm{T}}}{\sum_{j=1}^{N} D_j}\right], \tag{13}$$

where $\underline{e} = (1, \ldots, 1)^{\mathrm{T}}$. Consequently,

$$\tau_i^0 = \frac{hJ + \sum_{j=1}^{N}(\rho_i - \rho_j)D_j}{h\sum_{j=1}^{N} D_j} \qquad i = 1, \ldots, N-1 \tag{14}$$

Note that there are implicit lower and upper bounds on τ_i. An upper bound on τ_i, say τ_{u_i}, is obtained by allocating what is left after rationing lower priority customers, that is $J - \sum_{k<i} D_k\tau_k$, to customers of equal or higher priority; that is to customers $k \geqslant i$. Thus,

$$\tau_{u_i} = \frac{J - \sum_{k<i} D_k\tau_k}{\sum_{k\geqslant i} D_k} \tag{15}$$

A lower bound on τ_i, say τ_{l_i}, is obtained by allocating what is left after rationing lower priority customers, namely $J - \sum_{k<i} D_k\tau_k$, minus the total demand of higer priority customers, namely $T\sum_{k>i} D_k$, to customer type i. Therefore

$$\tau_{l_i} = \max\left(\frac{J - T\sum_{k>i}^{N} D_k - \sum_{k<i} D_k\tau_k}{D_i},\ 0\right) \tag{16}$$

Consequently, the problem becomes

$$\min_{\tau_1,\ldots,\tau_{N-1}} C(\tau_1, \ldots, \tau_{N-1})$$

subject to

$$\tau_{l_i} \leqslant \tau_i \leqslant \tau_{u_i} \qquad i = 1, \ldots, N-1$$

Since $\tau_{l_{k-1}}$ and $\tau_{u_{k-1}}$ are independent of τ_k, it follows that τ_{k-1} is independent of τ_k. However, since τ_{l_k} and τ_{u_k} are dependent on τ_{k-1}, it follows that τ_k is dependent on τ_{k-1} for all $k = 2, \ldots, N-1$. Consequently, we can first decide τ_1^* and after that we can decide τ_2^*, and so on. We also know that

$$\tau_1^0 = \frac{J}{\sum_{j=1}^{N} D_j} - \frac{\sum_{j=1}^{N}(\rho_j - \rho_1)D_j}{h\sum_{j=1}^{N} D_j} < \tau_{u_1} \tag{17}$$

Using the above information, we present a sequential procedure to find optimal rationing trigger times. The idea is simple; if the unrestricted minimum is in the range of the bounds, it is optimal. Else, either the lower bound or upper bound is optimal.

Sequential algorithm

Step 1

$$\tau_1^* = \begin{cases} \tau_{l_1} & \text{if } \tau_{u_1} - \tau_{l_1} \leqslant \dfrac{\sum_{j=1}^{N}(\rho_j - \rho_1)D_j}{h\sum_{j=1}^{N} D_j} \\ \tau_1^0 & \text{else} \end{cases}$$

Step k After we obtain $\tau_1^*, \ldots, \tau_{k-1}^*$, and τ_{l_k} and τ_{u_k}, we set

$$\tau_k^* = \begin{cases} \tau_{l_k} & \text{if } \tau_k^0 \leqslant \tau_{l_k} \\ \tau_k^0 & \text{if } \tau_{l_k} < \tau_k^0 < \tau_{u_k} \\ \tau_{u_k} & \text{if } \tau_k^0 \geqslant \tau_{l_k} \end{cases}$$

Note: If any $\tau_k^* = \tau_{l_k}$ and $\tau_{l_k} > 0$, then $\tau_j^* = T$ for all $j > k$.

Now we consider two extreme cases in terms of the holding cost. If there is no holding cost, the problem reduces to a Linear Program and clearly it is optimal to satisfy the demands of the highest customer first. If the holding cost becomes extremely large, there is only one feasible solution and it is not to ration. If the holding cost is positive but finite, the solution is a combination of these two extreme solutions.

Example 2 Let $\underline{D}^{\mathrm{T}} = (2, 3, 2)$, $\underline{\rho}^{\mathrm{T}} = (5, 10, 12)$, $h = 1$, $J = 50$, $T = 30$. We want to find optimal rationing trigger times using this data.

Step 1 $\tau_{l_1} = 0$, $\tau_{u_1} = \frac{50}{7}$. Since

$$\tau_{u_1} - \tau_{l_1} = \frac{50}{7} > \frac{\sum_{j=1}^{N}(\rho_j - \rho_1)D_j}{h\sum_{j=1}^{N} D_j} = \frac{29}{7},$$

$$\tau_1^* = \tau_1^0 = 3$$

That is, we start rejecting demands of customer type 1 at time 3.

Step 2 $\tau_{l_2} = 0, \tau_{u_2} = \frac{44}{5}$. Since $\tau_{l_2} < \tau_2^0 = 8 < \tau_{u_2}$, $\tau_2^* = \tau_2^0 = 8$. That is, we start rejecting demands of customer type 2 at time 8.

Obviously, we satisfy demands of customer type 3 until the end of the planning horizon or the inventory becomes 0, whichever comes first.

We now show that for sufficiently small changes in initial inventory, the optimal rationing trigger times change in the same proportion regardless of the lost sales costs and demand rates.

Property 2

$$\underline{\tau}^*(J + \Delta J) = \tau^*(J) + \text{constant}. \tag{18}$$

Proof If J is changed to $J + \Delta J$, then new $\underline{g}$ becomes

$$\underline{g}^{\text{new}} = \left[\frac{h(J + \Delta J)}{D_N} - \rho_N\right]\underline{D} + \text{diag}(\rho)\underline{D}$$

Then

$$\underline{\tau}^*(J + \Delta J) = H^{-1}\underline{g}^{\text{new}} = \underline{\tau}^*(J) + \frac{h\Delta J H^{-1}\underline{D}}{D_N}$$
$$= \underline{\tau}^*(J) + \frac{\Delta J \underline{e}}{\sum_{j=1}^{N} D_j}$$

Since $\Delta J\underline{e}/\sum_{j=1}^{N} D_j$ is independent of the decision variables, the proof is completed. □

A continuous review model with compound Poisson demand

In this section we assume that inventory levels are reviewed continuously. Moreover, demand follows a compound Poisson process. That is, the arrival process of customers follows a stationary Poisson process with rate λ, which can be decomposed into two independent Poisson processes with respective rates λ_1 and λ_2, and the demand from each customer is a random variable. We also make the following assumptions:

(a) Stock is replenished according to a (Q, r) policy. That is, when the inventory level reaches r, an order for Q units is placed.
(b) A lead time τ is a random variable.
(c) There is a single support level, say K, such that when the inventory level reaches K all low priority demands are backordered.
(d) $r > K > 0$.

Fill rates are used to compare the effects of rationing for this model. Let μ be the mean of the demand requested by an arriving customer. Since each cycle has expected cycle length $Q/(\lambda\mu)$, the expected number of backorders per unit time for type i customer is given by B_i/Q. So the fill rate for type i customer, say F_i is

$$1 - \frac{\lambda B_i}{\lambda_i Q}. \tag{19}$$

The system fill rate is simply

$$1 - \frac{B_1 + B_2}{Q}. \tag{20}$$

Note that as the support level K increases, F_1 increases (since B_1 decreases) and F_2 decreases. Also the fill rates increase as the order quantity increases.

We develop a simulation model under these assumptions since we cannot obtain analytical solutions for this case. See Appendix C for a flowchart for the simulation model. (We only provide a flowchart for the customer arrival routine for brevity). The model has been coded using Microsoft Visual Basic 5.0.

Example 3 There is an inventory system which uses a continuous review (Q, r) policy with a single support level. The arrival process of each type of customer follows a stationary Poisson process, and the demand requested by each customer is a Normal random variable with mean 10 and standard deviation 3. The lead time follows a uniform distribution with the range between 9 and 11 time units. The order quantity can be decided by the user, and here we arbitrarily chose $Q = 2r$.

The fill rates for a variety of system parameters which have been obtained by the simulation are presented in Table 3. We ran the simulation for 100 order cycles, and the warming-up period has been set to 5 cycles. From Table 3 we can observe that the fill rates tend to increase as the reorder point r increases. Also the fill rates for high priority customers are improved when the support level K increases.

We developed a diagram which can be designed to cover the range of parameter values observed in any particular system, so that both a support level and a reorder point could be chosen by the user to satisfy specified fill rate performance criteria. For example, Figure 2 shows the fill rates for each type of customer as the support level changes. (The data are $r = 400, \lambda_1 = 2, \lambda_2 = 3$.) Suppose the user wants to maintain $F_1 \geqslant 0.93$ and $F_2 \geqslant 0.80$, the support level can be chosen between 58 and 72.

Concluding remarks

We have developed both analytical and simulation models to improve the existing study. Firstly, we have extended the model with a single support level to the model with multiple support levels. Secondly, a simulation model whose assumption is more realistic than the one used in the

Table 3 Fill rates for a variety of system parameters

λ_1	λ_2	K	r	B_1	B_2	F_1	F_2	F
2	2	50	300	29.10	70.40	0.9030	0.7653	0.8342
2	2	100	300	14.68	95.07	0.9511	0.6831	0.8171
2	2	50	400	4.60	24.69	0.9885	0.9383	0.9634
2	2	100	400	1.55	46.06	0.9961	0.8849	0.9405
2	3	50	300	49.50	142.28	0.7937	0.6048	0.6804
2	3	100	300	32.50	175.23	0.8646	0.5133	0.6538
2	3	50	400	18.31	81.54	0.9428	0.8301	0.8752
2	3	100	400	6.37	112.08	0.9801	0.7665	0.8519

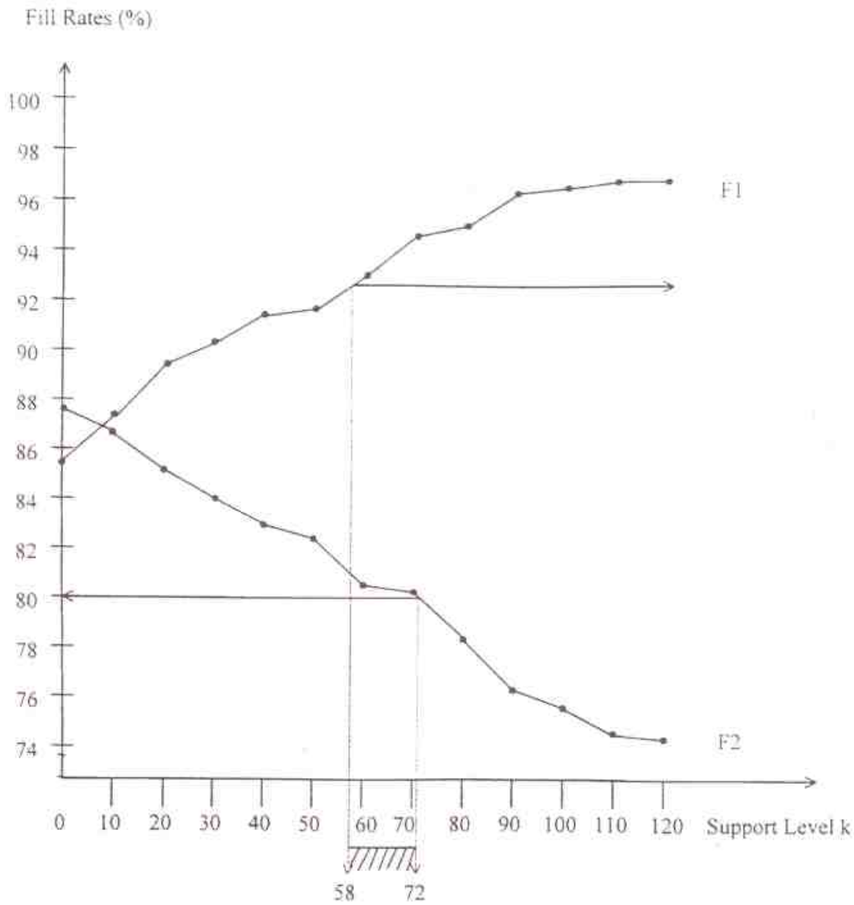


Figure 2 Fill rates for a variety of support levels.

analytical model has been presented. Thirdly, a single period optimisation model using the rationing trigger time concept has been developed. Finally, we have extended a model with a Poisson process demand to one with a compound Poisson process demand.

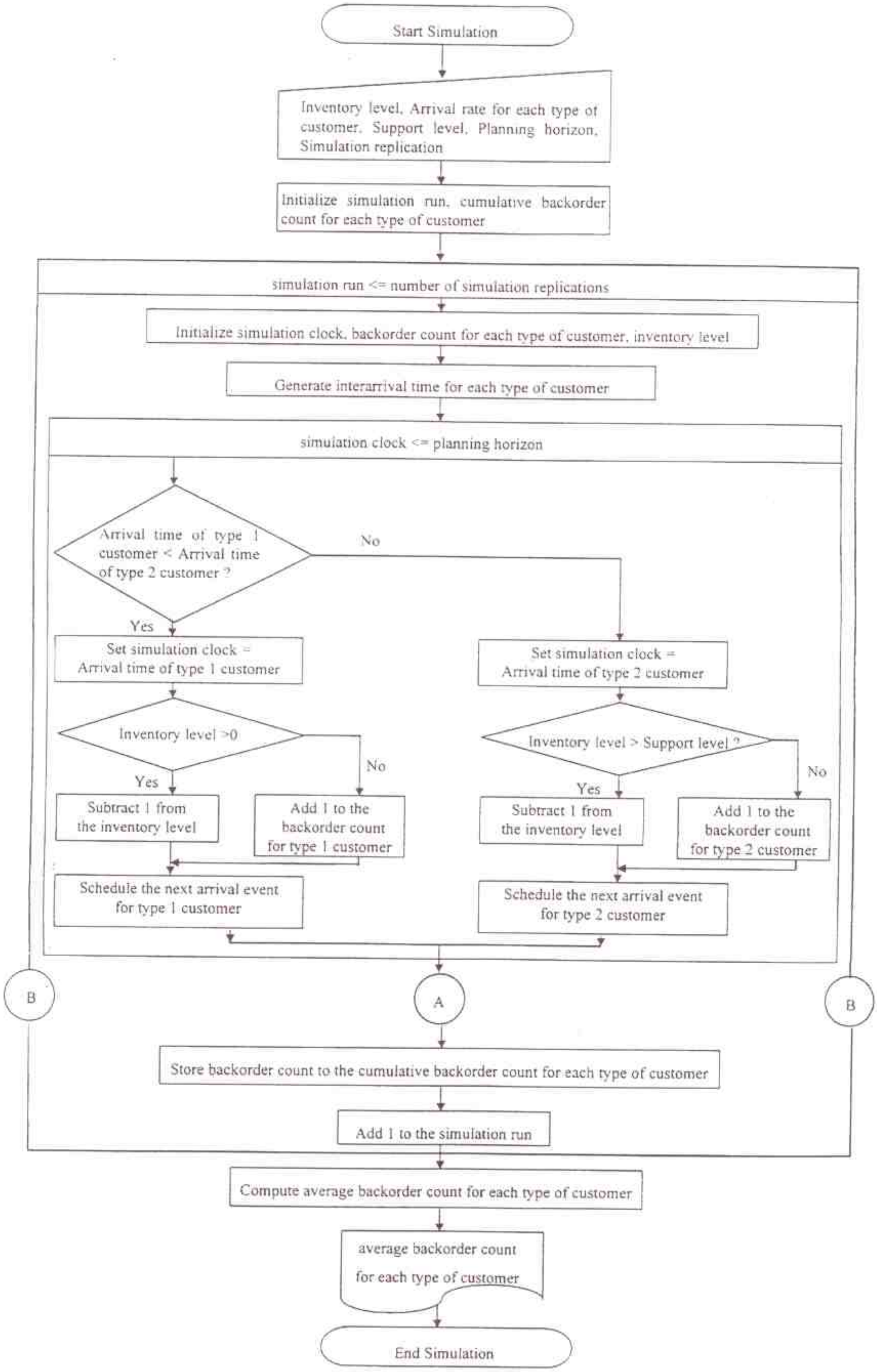
Decision makers may need to decide whether they choose an analytical model with less realistic assumptions or a simulation model with more realistic assumptions. Simulation is a useful technique for many complex inventory systems, such as those in this paper, which preclude any possibility of analytical solutions. Refer to Law and Kelton[8] for simulation models for some inventory systems.

Distinguishing between various priority classes of demand occurs in many areas such as airline and hotel reservation, military operations, hospital emergency rooms, computer storage allocation, sales company, etc. The results in this paper might be used as helpful managerial tools for these types of inventory/distribution systems. Case studies based on the results of this paper might be interesting research problems.

Acknowledgements—The authors are very grateful to the anonymous referees for their constructive comments. The authors are also grateful to Mr Jongsool Lee for writing the computer program and performing the computational experiments.

Appendix A

Flowchart for the simulation algorithm



Appendix B

Pseudo-code for the simulation algorithm

begin
Initialise simulation run, cumulative backorder count for each type of customer
While simulation run $\leqslant$ number of simulation replications **do**
T := 0;
backorder count for each type of customer := 0;
inventory level;
Generate A1, A2
While $T \leqslant$ planning horizon **do**
If $A1 < A2$
then T := A1
  if inventory level > 0
  then inventory level := inventory level − 1;
  else backorder count for type 1 customer := backorder count for type 1 customer + 1;
  Schedule A1; {arrival time of type 1 customer}
else T := A2
  if inventory level > support level
  then inventory level := inventory level − 1;
  else backorder count for type 2 customer := backorder count for type 2 customer + 1;
  Schedule A2; {arrival time of type 2 customer}
end_while
cumulative backorder count := cumulative backorder count + backorder count for each type of customer; simulation run := simulation run + 1;
end_while
Compute average backorder count for each type of customer;
end