An Embedded Markov Chain Approach to Stock Rationing

Mehmet Murat Fadıloğlu[a], Önder Bulut[b]

[a]*Department of Industrial Systems Engineering, Izmir University of Economics, Izmir, Turkey*

murat.fadiloglu@ieu.edu.tr

[b]*Department of Industrial Engineering, Yaşar University, Izmir, Turkey*

onder.bulut@yasar.edu.tr

September, 2010

We propose a new method for the analysis of lot-per-lot inventory systems with backorders under rationing. We introduce an embedded Markov chain that approximates the state-transition probabilities. We provide a recursive procedure to generate these probabilities and obtain the steady-state distribution.

Keywords: Inventory models; Rationing; Embedded Markov chains

---

1. Introduction

Rationing is an inventory policy that allows prioritization for different demand classes. This prioritization results in a different service level for each demand class. The mechanism through which the rationing policy is implemented is to stop serving lower priority classes when the inventory on hand drops below a certain critical level. Under this level, only higher priority classes are served. The same critical level can also be used for clearing backorders. This clearing mechanism is called priority clearing.

In this paper, we present a new method for the analysis of continuous-review inventory systems with backordering under $(S-1, S)$ ordering policy and critical level stock-rationing with priority clearing. In the system of interest, every time a demand occurs, a replenishment order is placed, which arrives after a deterministic lead-time. If the arriving demand belongs to class 1, the demand is instantaneously satisfied if there is any inventory on-hand. Otherwise, the class 1 demand is backordered.

Rationing only affects class 2 demands. The rationing policy is specified by a support level $K \le S$. If the inventory level is above K, arriving class 2 demands are instantaneously satisfied, otherwise they are backordered. The support level K also determines how class 2 backorders are cleared. At any inventory level except K, replenishment orders are used to increase the inventory level. When a replenishment order arrives and finds the inventory level at K, then that order is used to clear the oldest class 2 backorder unless there is none. Thus, the inventory level cannot exceed the support level unless all backorders are cleared. This clearing is called priority clearing in the literature. Although this clearing mechanism is an integral part of the rationing policy (when applied to inventory systems with backordering), other clearing mechanisms have also been suggested due to the fact that an exact analysis of rationing with priority clearing is not available.

This paper provides an algorithm to approximate the steady-state distribution for the inventory system of interest. Once the steady-state distribution is obtained, it can be used to estimate any long-run performance measure. Our method is based on the observation that the state of the inventory system sampled at multiples of the supply lead-time evolves according to a Markov chain. Although there are many approximate results in the literature on this important inventory system, there was no method capturing priority clearing dynamics. This constitutes the main contribution of the paper. Our approach takes into account that the dynamics of the rationing policy not only depends on the number of arrivals of different types, but on their relative ordering as well. The effect of the relative ordering is quite complex as can be observed in the recursive procedure we propose. Hence the problem we are attacking is combinatorial in nature. This explains why the problem resisted many attempts for its solution.

Nahmias and Demmy [12] are the first to consider a continuous-review environment with multiple demand classes and backorders. They derive approximate expressions for the expected number of backorders and for the fill rates. Deshpande et al. [6] and Arslan et al. [1] study the problem in the same setting and derive approximate results. Their analyses yield the same results with Dekker et al. [4] under the lot-for-lot policy. Kocaga and Sen [9] extend the approximation of Dekker et al. [4] to accommodate a demand lead-time. The aforementioned literature either ignores the clearing issue ([12], [4], and [9]) or resorts to tractable clearing mechanisms that eliminate interaction between consecutive lead-time periods ([6], [1]). Deshpande et al. [6] address this issue on page 684 of their study: "The optimal scheme is to always clear higher-priority customers first. However, this "priority-clearing" scheme is intractable because it does not allow closed-form expressions for the stockout levels, and average number of demand in backlog, for each demand class. To overcome this problem we introduce a tractable “threshold clearing” scheme to approximate the systems dynamics”. Our algorithm provides an improvement over these approaches as discussed in Section 3.

The clearing issue is not relevant for the lost sales case. Under lost sales, Dekker et al. [5] provide exact service levels for $(S-1,S)$ policy, while Melchiors et al. [11] provide an approximate analysis for (Q,R) policy. There are also few works in the literature that consider dynamic rationing policies ([7], [10], and [14]).

2. The Embedded Markov Chain

We assume inventory for an item is held and replenished over time in order to keep up with the demand from two customer classes, which occur according to two independent Poisson processes with rates λ_1 and λ_2, accordingly. That is, the total demand also follows a Poisson process with rate $\lambda = \lambda_1 + \lambda_2$. Replenishment orders arrive after a deterministic lead-time $L > 0$. The inventory policy is $(S-1,S)$ with static rationing. Any unmet demand is backordered and they are cleared according to the priority clearing mechanism.

We define the state of the system at time t as $(X(t),B(t))$, where *X(t)* is the number of outstanding replenishment orders at time *t,* and *B(t)* is the number of class 2 backorders at time *t*. Under a lot-for-lot inventory policy, *X(t)* is also equal to the number of demand arrivals of both types in $(t-L,t]$, since each demand arrival triggers a replenishment order. Each demand occurrence either decreases the inventory level or causes a class 2 backorder. Thus, the inventory level at time *t* is the inventory position at time $t-L$, which is *S*, minus the effect of demand on inventory in $(t-L,t]$, which is $X(t)-B(t)$, i.e.,

$$I(t) = S - X(t) + B(t). \tag{1}$$

There is no backorder if the inventory level is above *K*, since all backorders need to be cleared at the support level, i.e.,

$$B(t) = 0 \text{ if } X(t) < S - K . \tag{2}$$

Only the class 2 demands that arrive after the inventory level hits K are backordered. Even if all the demand arrivals in $(t-L,t]$ belong to class 2, the maximum number of backorders would be $X(t)-(S-K)$. Thus,

$$0 \le B(t) \le X(t)-(S-K) \quad \text{if } X(t) \ge S-K\,. \tag{3}$$

Conditions (2) and (3) specify the feasible states and thereby the state space of the embedded Markov chain, while Equation (1) specifies the inventory level corresponding to each state of the state space.

Given that we know the state of the system at time t, it is possible to derive the probability that the system reaches a certain state at time t+L. If we derive this probability for all feasible states at time t, and at time t+L, then we obtain an embedded Markov chain for the inventory system. These probabilities are the one-step transition probabilities of the Markov chain. They determine the probabilistic evolution of the inventory system at multiples of lead-time. One should note that the original continuous-time process describing the evolution of the states at any point in time is regenerative. The process regenerates itself every time there is no outstanding replenishment order in the inventory system, i.e., when $\left(X(t),B(t)\right)=(0,0)$. Since the underlying continuous-time process is regenerative, the process is ergodic and the limiting distribution of the process exists (See Stidham [13]). Therefore, for any ε value there exist a t_ε such that

$$\left|P\left\{X(t)=x,B(t)=b\right\}-P\left\{X=x,B=b\right\}\right| \le \varepsilon \quad \text{for all} \quad t \ge t_\varepsilon\,.$$

This means that

$$\left|P\left\{X(nL)=x,B(nL)=b\right\}-P\left\{X=x,B=b\right\}\right| \le \varepsilon \quad \text{for all} \quad n \ge \left\lceil \frac{t_\varepsilon}{L} \right\rceil.$$

Hence, the limiting distributions of the continuous-time process and the embedded Markov chain are the same, i.e., the limiting distribution of the embedded Markov chain is sufficient for statistical characterization of the inventory system.

One of the state variables, *X(t),* sampled at multiples of lead-time, evolves on its own according to an embedded Markov chain. Moreover, the one-step transition probabilities of this embedded Markov chain are independent of the origin state, i.e.,

$$P\left\{X(t+L)=x_L \mid X(t)=x_0\right\}=P\left\{D_{(t,t+L]}=x_L\right\}=e^{-\lambda L}\frac{(\lambda L)^{x_L}}{x_L!}\ \ for\, x_L=0,1,2,\dots\ . \tag{4}$$

The evolution of *X(t)*, number of outstanding replenishment orders, is fully independent of the rationing policy. The result stated in (4) is the basis for the steady-state analysis of $(S-1,S)$ inventory systems under constant lead-times (see Hadley and Whitin [8] pages 204-205). Our approach can be considered as an extension of that classical analysis. A direct implication of (4) is that the distribution of *X(t)* converges to its limiting distribution at $t=L$. Thus, we are able to decouple one of the dimensions of the two-dimensional chain, and analyze it independently. This simplifies our approach considerably.

We can express all first step probabilities as $P\left\{X(t+L)=x_L,B(t+L)=b_L \mid X(t)=x_0,B(t)=b_0\right\}$. The probabilities that relate to the event of reaching an inventory level above *K* can be obtained directly from (4) as

$$P\left\{X(t+L)=x_L,B(t+L)=0 \mid X(t)=x_0,B(t)=b_0\right\}=P\left\{D_{(t,t+L]}=x_L\right\}=e^{-\lambda L}\frac{(\lambda L)^{x_L}}{x_L!}\ ,$$
$$\text{for } 0\le x_L\le S-K, \text{ for all feasible } (x_0,b_0) \text{ pairs,}$$

since there will be no backorders when the inventory level is above the support level.

One needs considerably more effort in order to obtain other one-step transition probabilities. Since we know the distribution of *X(t+L)* and the fact that the distribution is independent of *X(t)* and *B(t),* we can use this for our objective in

$$P\left\{X(t+L)=x_L,B(t+L)=b_L \mid X(t)=x_0,B(t)=b_0\right\}=$$
$$P\left\{B(t+L)=b_L \mid X(t+L)=x_L,X(t)=x_0,B(t)=b_0\right\}e^{-\lambda L}\frac{(\lambda L)^{x_L}}{x_L!}\ . \tag{5}$$

Hence, we need to compute

$$P\left\{B\left(t+L\right)=b_L \mid X(t+L)=x_L, X(t)=x_0, B\left(t\right)=b_0\right\} \tag{6}$$

for all feasible state pairs $\left(x_0, b_0\right)$ and $\left(x_L, b_L\right)$. This is the probability that there are b_L units of backorder at time $t+L$, given that there are b_0 units of backorder at time t, x_0 demand arrival occurs in $(t-L,t]$, and x_L demand arrivals occurs in $(t,t+L]$. The reader should note that x_0 is also the number of replenishment arrivals in $(t,t+L]$.

Since the number of order and replenishment arrivals during $(t,t+L]$ is known, it is the order of the replenishment and demand arrivals, and the class of the demand arrivals that determines the number of backorders at time $t+L$. Unfortunately, it is not possible to obtain a closed form expression for (6). Yet, it is still possible to compute these probabilities using a recursive procedure. This procedure is based on a related probability expression, which is

$$P\left\{B\left(t+L\right)=b_L \mid X\left(t+L\right)=x_L, B\left(t'\right)=b, Y\left(t'\right)=y, Z\left(t'\right)=z\right\} \quad \text{for } t \le t' \le t+L \tag{7}$$

where $Y(t')$ is the number of demand arrivals in $(t',t+L]$, and $Z(t')$ is the number of replenishment arrivals in $(t',t+L]$. Here we have to use t' instead of t, since we need probabilities on what happens between any point in time -- within the lead-time period -- and the end of the period. The reader should note that (6) can be expressed in terms of (7) as

$$\begin{aligned} &P\left\{B\left(t+L\right)=b_L \mid X\left(t+L\right)=x_L, X\left(t\right)=x_0, B\left(t\right)=b_0\right\}= \\ &\qquad P\left\{B\left(t+L\right)=b_L \mid X\left(t+L\right)=x_L, B\left(t\right)=b_0, Y\left(t\right)=x_L, Z\left(t\right)=x_0\right\}. \end{aligned} \tag{8}$$

In order to devise a method to compute the probabilities of (7), we condition them on the first arrival. This can either be a demand arrival or a replenishment arrival. We define $f(u \mid b,y,z,t')$ as the density function for the first arrival time after t', given that there are b units of backorder at time t', and the number of demand and replenishment arrivals to occur in $(t',t+L]$ are y and z, respectively. We also define $P\left\{A(u)=a \mid b,y,z,t'\right\}$ where $a \in \left\{d,r\right\}$ as the probability that the first occurrence, which takes place at time *u*, is a replenishment or a demand arrival under the same conditions.

PROPOSITION 1. Let $p_1 = \lambda_1/(\lambda_1+\lambda_2)$, $p_2 = 1-p_1$, $x = x_L + z - y$, $y \geq 0$, $z \geq 0$, and $b \in \left[0,\left(x-(S-K)\right)^+\right]$. The following recursive equation holds for $\max(y,z) \neq 0$:

$$
\begin{aligned}
&P\{B(t+L)=b_L \mid X(t+L)=x_L, B(t')=b, Y(t')=y, Z(t')=z\} = \\
&\int_{u=t'}^{u=t+L}\Bigg[P\{A(u)=r \mid b,y,z,t'\}\,P\left\{B(t+L)=b_L \,\middle|\, \begin{array}{c} X(t+L)=x_L, B(u)=\left(b-\mathbf{1}\{x-b=S-K\}\right)^+, \\ Y(u)=y, Z(u)=(z-1)^+ \end{array}\right\} \\
&\quad + P\{A(u)=d \mid b,y,z,t'\}\Bigg(p_1 P\left\{B(t+L)=b_L \,\middle|\, \begin{array}{c} X(t+L)=x_L, B(u)=b, \\ Y(u)=(y-1)^+, Z(u)=z \end{array}\right\} \\
&\quad + p_2 P\left\{B(t+L)=b_L \,\middle|\, \begin{array}{c} X(t+L)=x_L, B(u)=b+\mathbf{1}\{x-b\geq S-K\}, \\ Y(u)=(y-1)^+, Z(u)=z \end{array}\right\}\Bigg)\Bigg] f(u \mid b,y,z,t')\,du
\end{aligned}
\tag{9}
$$

Moreover, the following boundary condition holds:

$$
P\{B(t+L)=b_L \mid X(t+L)=x_L, B(t')=b, Y(t')=0, Z(t')=0\} = \mathbf{1}\{b_L=b\} \tag{10}
$$

PROOF: Given $X(t+L)=x_L, Y(t')=y, Z(t')=z$, the number of outstanding replenishment orders at time *t'* is determined by

$$
X(t') = X(t+L) + Z(t') - Y(t') = x_L + z - y. \tag{11}
$$

The logic behind Equation (11) can be explained as follows: In order to find the number of outstanding replenishment orders at $X(t+L)$, one should add the difference between the number of demand arrivals in $(t',t+L]$ (which trigger new replenishment orders) and the number of replenishment arrivals in $(t',t+L]$ (which clear the outstanding replenishment orders) to the number of outstanding replenishment orders at *t'*. When the value of $X(t')$ is denoted as *x*, Inequalities (2) and (3) yield the condition $b \in \left[0, \left(x-(S-K)\right)^{+}\right]$.

The proposition is based on the law of total probability. The probability $P\left\{B(t+L)=b_L \mid X(t+L)=x_L, B(t)=b_0, Y(t)=x_L, Z(t)=x_0\right\}$ is conditioned on the time and the type of the first arrival in $(t',t+L]$. There are three distinct types of arrival that can take place: a replenishment order arrival, a class 1 demand arrival, and a class 2 demand arrival. The probability that the first arrival in $(t',t+L]$ is at time *u* and is a replenishment order arrival is $P\left\{A(u)=r \mid b,y,z,t'\right\}.f\left(u \mid b,y,z,t'\right)du$, and the probability that the first arrival in $(t',t+L]$ is at time *u* and is a demand arrival is $P\left\{A(u)=d \mid b,y,z,t'\right\}.f\left(u \mid b,y,z,t'\right)du$. This demand arrival at time *u* is of class 1 with probability p_1, or of class 2 with probability $p_2 = 1-p_1$.

The effect of this first arrival depends on the number of outstanding replenishment orders (*x*) at that time due to the rationing policy. When a replenishment order arrives first, the number of replenishments to arrive is decreased by one and the number of class 2 backorders is decreased by one only if the system is at the clearing position, i.e., $x-b=S-K$, and if there is a backorder to be cleared. When a demand of class 1 arrives first, the number of demands to arrive is decreased by one. No other change is required since the decrease in inventory level is automatic via (11). Finally, when a demand of class 2 arrives first, the number of demands to arrive is again decreased by one, and the number of class 2 backorders is increased if the inventory level is at or below the support level, i.e., $x-b \geq S-K$.

Equation (10) determines the probabilities corresponding to the case at which no demand or replenishment arrival is left. In this case, the class 2 backorder level at the end the period $(t',t+L]$ is the level at *t'*, since there is no arrival left to change the state of the system. □

Proposition 1 expresses probabilities in (7) as a function of probabilities of the same type, which correspond to smaller number of arrivals. Note that these probabilities depend on *t'*, which means that one would have to compute a different probability for every instant in $(t,t+L]$. However, if the replenishment and the demand arrivals in $(t,t+L]$ are uniformly distributed, these probabilities are independent of *t'*. In this case, the probabilities only depend on the numbers of arrivals which are already specified.

PROPOSITION 2. Given the number of demand arrivals in $(t-L,t]$ and in $(t,t+L]$, the replenishment times and the demand arrival times in $(t,t+L]$ are iid random variables with uniform distribution over $(t,t+L]$.

PROOF. Since arrivals occur according to a Poisson process, the unordered arrival times in $(t-L,t]$ are independent random variables with uniform distribution on $(t-L,t]$ and the unordered arrival times in $(t,t+L]$ are independent random variables with uniform distribution on $(t,t+L]$. Each demand arrival in $(t-L,t]$ triggers a replenishment order that arrives exactly in L units of time. This means that the replenishment order arrival times in $(t,t+L]$ are just a translation of the demand arrivals in $(t-L,t]$ by L units. Hence, the unordered replenishment order arrival times in $(t,t+L]$ are independent random variables with uniform distribution on $(t,t+L]$. Moreover, since demand arrival times in $(t-L,t]$ and $(t,t+L]$ are independent, the demand and the replenishment arrival times in $(t,t+L]$ are all independent. □

Based on Proposition 2, one can conclude that the arrival times are uniformly distributed. However, this is not true. The probabilities in (6) not only specify the number of arrivals to occur, but also the number of class 2 backorders. The class 2 backorders at t are dependent on the demand times in $(t-L,t]$, which means that under the conditions that are given in (6), the replenishment arrivals in $(t,t+L]$ are no longer uniformly distributed. But we claim that if that dependence is ignored and the uniformity —which is valid without the knowledge of backorders—is assumed, the resulting analysis is an excellent approximation. We demonstrate the quality of the approximation in Section 3.

Once the uniformity on $(t,t+L]$ is assumed, we know that the remaining arrivals on $(t',t+L]$ are still uniformly distributed for any $t'\in(t,t+L]$. Thereby,

$$P(A(u)=r\,|\,b,y,z,t')=\frac{z}{y+z}\ ,\ P(A(u)=d\,|\,b,y,z,t')=\frac{y}{y+z},$$

and all the probabilities in (9) are independent of the first arrival time, u. When these are applied on Equation (9), we obtain

$$
\begin{aligned}
&P\left\{B(t+L)=b_L \mid X(t+L)=x_L, B(t')=b, Y(t')=y, Z(t')=z\right\}= \\
&\qquad \frac{z}{z+y} P\left\{B(t+L)=b_L \left| \begin{array}{c} X(t+L)=x_L, B(t')=\left(b-\mathbf{1}\{x-b=S-K\}\right)^+, \\ Y(t')=y, Z(t')=(z-1)^+ \end{array} \right.\right\} \\
&\qquad +\frac{y}{z+y}\left(p_1 P\left\{B(t+L)=b_L \left| \begin{array}{c} X(t+L)=x_L, B(t')=b, \\ Y(t')=(y-1)^+, Z(t')=z \end{array} \right.\right\}\right. \\
&\qquad\qquad \left. + p_2 P\left\{B(t+L)=b_L \left| \begin{array}{c} X(t+L)=x_L, B(t')=b+\mathbf{1}\{x-b\geq S-K\}, \\ Y(t')=(y-1)^+, Z(t')=z \end{array} \right.\right\}\right)
\end{aligned} \quad . \quad (12)
$$

Equation (12) is independent of t'. Thereby, it provides us with a practical means to calculate probabilities of (6).

The reader should note that if we know the probabilities corresponding to fewer arrivals, then we can compute the probabilities corresponding to more arrivals via (12). This fact constitutes the basis of the recursive procedure we propose. All that is needed to complete the procedure is Equation (10), which determines the probabilities corresponding to the case at which no demand or replenishment arrival is left. Starting with this probability, all other probabilities of interest are computed in a recursive fashion, invoking the interior equation (9) and increasing $Y(t')$ and then $Z(t')$ one by one. Finally, by applying (8) and then (5), we obtain the one-step transition probabilities.

All the one-step transition probabilities can be generated with this procedure. The only problem is that the state space of the embedded Markov chain is infinite and it would take infinite amount of time to generate all the elements of the Markov transition matrix. Thereby, we have to find a way of working with a finite version, i.e., a truncation, of the original matrix. This problem is resolved in Section 3.

3. Steady-State Analysis

We consider a subset of the state space for which $0 \leq X(t) \leq D_{\max}$. $D_{\max}$ is the maximum number of outstanding replenishment orders we consider. Since the second dimension of the state space, the number of class 2 backorders, is limited by the number of outstanding replenishment orders through (2) and (3), no truncation is needed for this dimension, i.e., $B(t) \le D_{\max} - (S - K)$. One should note that when we truncate the state-space as described, we are effectively ignoring an infinite number of states whose total probability is equal to $P\{X(t) > D_{\max}\}$. Since *X(t)* has a Poisson distribution with parameter λL as stated in (4), we are ignoring the tail of Poisson distribution. Thereby, as $D_{\max}$ increases, the probability corresponding to the ignored part of the state space goes to zero rapidly.

In order to perform steady-state analysis, the transition matrix of the embedded Markov chain is needed. Since it is not possible to obtain the full transition matrix, we generate a submatrix which corresponds to the states that are conserved by the truncation, i.e., that are not ignored. We call this submatrix **Q**. The number of states corresponding to **Q** is $(S-K) + (D_{\max} - (S-K) + 1)(D_{\max} - (S-K) + 2)/2$. When $X(t) < (S - K)$, then $B(t) = 0$, which means there are exactly $(S-K)$ states corresponding to that part of the state space. When $X(t) \ge (S - K)$, then the number of feasible $B(t)$ values increases one by one starting from one for $X(t) = (S - K)$. Then total number of states in this part of the state space forms an arithmetic progression, which is $(D_{\max} - (S-K) + 1)(D_{\max} - (S-K) + 2)/2$ at $X(t) = D_{\max}$.

The number of states increases with $D_{\max}$ and is in the order of $D_{\max}^2$. The recursive procedure discussed in the previous section computes the probability given in expression (7), for all feasible x_L, b_L, b, y, and z values. Each of these dimensions is bounded by $D_{\max}$, since our truncation neglects the event that the number of demand or replenishment arrivals during lead-time is greater than $D_{\max}$. This means the computational complexity of the recursive algorithm is $O\left(D_{\max}^5\right)$. Thus, as a result of a computational effort of $O\left(D_{\max}^5\right)$, we are able to obtain the matrix **Q**, which is a finite submatrix of the original Markov transition matrix, which is infinite.

Since the whole transition matrix is not generated, an analysis based on this truncation would be an approximation. Although this claim is correct in the strictest sense, there is a theory in computational linear algebra, which enables us to get exact upper and lower bounds for steady-state probabilities corresponding to the nontruncated states by considering a truncated version of an irreducible Markov chain. The reader is referred for the theory behind the procedure giving the bounds to the seminal work by Courtois and Semal [2]. The upper and lower bounds for steady-state probabilities are given explicitly in Dayar and Stewart [3]. Although the theory behind the bounds is not new, its application to the analysis of infinite state-space inventory systems is. These bounds, which are known to be loose in a general setting, work extremely well for the inventory system under consideration. It is observed that the upper and lower bounds for our problem converge together rapidly as $D_{\max}$ is increased. Thus, we can get the steady-state probabilities with any desired accuracy.

In our problem, the size of **Q** is determined by the maximum demand considered during lead-time, $D_{\max}$. The dimension of **Q** is equal to the number of states considered, and it is in the order of $D_{\max}^2$. Since one needs to invert **Q** to obtain the aforementioned bounds, these bounds can be obtained with a computational effort of $O\left(D_{\max}^6\right)$. The recursive algorithm generating the matrix **Q**, has a computational complexity of $O\left(D_{\max}^5\right)$ as we have discussed before. Thus, the computational complexity of the whole procedure is $O\left(D_{\max}^6\right)$. That is as $D_{\max}$ increases, the computation times increases accordingly. However, as $D_{\max}$ increases, a greater portion of the system's dynamics is included in the matrix **Q**, i.e., bounds become tighter.

4. Numerical Study

In order to demonstrate the quality of the approximation involved in our approach, we provide the results of a numerical experiment. We generate different system settings varying in policy parameters and arrival rates. In Table 1, we provide the estimates for fill rates of the first and second classes (β_1, β_2) and the respective computation times. The algorithm described in this paper was coded in MATLAB™ computing language, then converted to C using MATLAB C Compiler, and was run on a PC with an AMD Athlon processor of 2.5 GHz clock speed and 1 GB of RAM. Since the exact steady-state probabilities are not computed, but upper and lower bounds are; their estimate, $\hat{P}\left\{X = x, B = b\right\}$, is defined to be the midpoint of the bounds. The probabilities of the states ignored by the truncation are assumed to be zero. Hence, using PASTA property for Poisson arrivals we obtain

$$\hat{\beta}_1 = \sum_{x=0}^{D_{\max}} \sum_{b=\left(x-(S-1)\right)^+}^{x-(S-K)} \hat{P}\left\{X = x, B = b\right\}$$

$$\hat{\beta}_2 = \sum_{x=0}^{\min\{S-K-1, D_{\max}\}} \hat{P}\{X = x, B = 0\} .$$

It should be also noted that there is no need to calculate the fill rate of the second class via our procedure. Since the steady-state distribution for the random variable *X(t)* is Poisson as stated in (4), we know that

$$\beta_2 = \sum_{x=0}^{S-K-1} P\{X = x, B = 0\} = \sum_{x=0}^{S-K-1} P\{X = x\} = \sum_{x=0}^{S-K-1} e^{-\lambda L} \frac{(\lambda L)^x}{x!} . \qquad (13)$$

Using (13), we can compute the exact fill rate for the second class. Nevertheless, we also provide our estimate for β_2 in order to demonstrate that our estimate converges to the theoretical value.

In order to obtain the true service levels for the system, we also developed a simulation code in C. This program was run for 6 million customers to make sure that the class 2 service measures obtained were exact to the third significant digit. We observed that the estimates given by our algorithm are stable for different $D_{\max}$ values depending on the expected number of arrivals during lead-time (λL) of the experiment. For low (λL = 1), medium (λL = 3), and high (λL = 6) arrivals, we used $D_{\max}$ values of 7, 13, and 18, respectively. The largest deviation of our algorithm from the simulation values is 0.007, which is observed in class 1 fill rate estimate when $S = 4,\ K = 2,\ \lambda_1 = 4.5,\ \lambda_2 = 1.5$. The average absolute deviation from the exact values across all the settings is 0.001. Hence, our algorithm results in high quality fill rate estimates. The approximation due to ignoring the dependence of the replenishment times with the class 2 backorder –discussed in Section 2– does not cause any significant deviation from the true service levels.

In Table 1, we also report the results of the analysis provided by Deshpande et al. [6]. This also compares our results with the ones of Dekker et al. [4] and Arslan et al. [1] since their results are the same with the one of Deshpande et al. [6] (the reader can find the equivalence proofs in [1] and [9]). We do not report their results for class 2 fill rate since they are exact like ours. For Deshpande et al.'s method, the largest deviation from the simulation result is 0.372 ($S = 4,\ K = 2,\ \lambda_1 L = 3,\ \lambda_2 L = 3$). It is interesting that although their maximum deviation is approximately 53 times larger compared to ours, they both occur at $S = 4,\ K = 2,\ \lambda L = 6$. The average absolute deviation in their case is 0.111, which is 111 times greater than ours. Furthermore, our algorithm is superior to theirs in every single setting considered in our study. It is also observed from Table 1 that the computation times for our algorithm increases with total demand rate. But, for all reasonable systems our method yields accurate results in a fraction of a minute.

Table 1 Comparison of Results with Simulation

			Our Algorithm			Deshpande's	Simulation	
			Time (sec)	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_1$	$\hat{\beta}_2$
$(S=4,\ K=1)$	$\lambda L = 1$	$p_1 =$ 0.25	0.1	0.995	0.920	0.994	0.995	0.920
		0.50		0.991	0.920	0.990	0.991	0.920
		0.75		0.986	0.920	0.985	0.986	0.920
	$\lambda L = 3$	$p_1 =$ 0.25	2.8	0.912	0.423	0.861	0.912	0.423
		0.50		0.824	0.423	0.764	0.824	0.423
		0.75		0.737	0.423	0.696	0.736	0.423
	$\lambda L = 6$	$p_1 =$ 0.25	15	0.787	0.062	0.499	0.788	0.062
		0.50		0.581	0.062	0.292	0.580	0.062
		0.75		0.381	0.062	0.198	0.376	0.062
$(S=4,\ K=2)$	$\lambda L = 1$	$p_1 =$ 0.25	0.2	0.999	0.736	0.998	0.999	0.736
		0.50		0.995	0.736	0.994	0.995	0.735
		0.75		0.989	0.736	0.988	0.989	0.736
	$\lambda L = 3$	$p_1 =$ 0.25	3.5	0.978	0.199	0.947	0.978	0.199
		0.50		0.913	0.199	0.845	0.913	0.200
		0.75		0.807	0.199	0.739	0.806	0.199
	$\lambda L = 6$	$p_1 =$ 0.25	17	0.947	0.017	0.730	0.948	0.017
		0.50		0.799	0.017	0.426	0.798	0.017
		0.75		0.569	0.017	0.246	0.562	0.017

Table 2 Impact of Increase in Total Demand Rate (S = 4, K = 2, p_1 = 0.75)

	$\hat{\beta}_1$		
λL	Our Algorithm	Deshpande's	Simulation
1	0.989	0.988	0.989
3	0.807	0.739	0.806
6	0.569	0.246	0.562
8	0.520	0.091	0.512
10	0.500	0.030	0.491
12	0.487	0.009	0.481

As shown in Table 2, increasing expected lead-time demand is detrimental for other approaches, since they all assume artificial clearing mechanisms that eliminate interaction between consecutive lead-time periods. As lead-time demand increases, this interaction gets stronger and, thereby, their approximation quality deteriorates. In our approach, the same effect is also present but to a significantly lesser extent, since we incorporate the interaction into the embedded Markov chain. Our approach is based on the –assumed– uniform distribution of the replenishment opportunity times given the information vector that consists of the number of replenishment opportunities and class 2 backorders. As total demand rate increases, the number of the class 2 backorders increases that adversely affects the validity of the uniformity assumption. But, the scope of this negative effect is limited and the degradation it causes does not exceed a certain level as total demand rate continues increasing as seen in Table 2. The effect of increasing the rationing level is also similar since it also causes more class 2 backorders and more interaction between lead-time periods. To summarize the above discussion, the results of all approaches degrade –although to different extents– as the expected lead-time demand and the rationing level increase. Our approach is significantly more robust compared to others.

Table 3 Impact of Increase in *S* and *K* (λ = 20, p_1 = 0.5)

		Our Algorithm			Simulation	
S	K	Time (sec)	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
10	2	69	0.749	0.001	0.753	0.001
	4	81	0.935	0.000	0.940	0.000
20	4	27	0.964	0.156	0.967	0.156
	8	43	0.995	0.021	0.998	0.021
30	6	8.8	0.999	0.786	0.999	0.786
	12	20	0.999	0.296	0.999	0.296
40	8	1.9	1.000	0.991	1.000	0.991
	16	8.6	1.000	0.786	1.000	0.786

Even though lot-per-lot policy is primarily for slow moving items, in Table 3 we report settings with a high demand rate in order to demonstrate the computation time and the accuracy of our algorithm for larger *S* and *K* values (*K* is set to one fifth and two fifths of *S*). For the constant $D_{\max}$ of 40 used in our experiment, computation times do not increase with *S*. Furthermore, it decreases as *S*-*K* increases, as suggested by the discussion of Section 3 on the size of the state-space. Just as in Tables 1 and 2, our algorithm results in high quality estimates which are exact to the second significant digit for all cases.

References

[1] H. Arslan, S.C. Graves, T. Roemer, A single-product inventory model for multiple demand classes, *Management Science* 53 (2007) 1486-1500.

[2] P.-J Courtois, P. Semal., Bounds for the positive eigenvectors of nonnegative matrices and for their approximations by decomposition, *Journal of the Association for Computing Machinery* 31 (1984) 826-838.

[3] T. Dayar, W.J. Stewart, Quasi lumpability, lower-bounding coupling matrices, and nearly completely decomposable Markov chains, *Siam J. Matrix Anal. Appl.* 18 (1997) 482-498.

[4] R. Dekker, M.J. Kleijn, P.J. de Rooij, A spare parts stocking policy based on equipment criticality, *Int. J. Production Economics* 56-57 (1998) 69-77.

[5] R. Dekker, R.M. Hill, M.j. Kleijn, R.H. Teunter, On the (S -1, S) lost sales inventory model with priority demand classes, *Naval Research Logistics* 49 (2002) 593-610.

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

[7] M.M. Fadıloğlu, Ö. Bulut, A dynamic rationing policy for continuous-review inventory systems, *European Journal of Operational Research* 202 (2010) 675-685.

[8] G. Hadley, T. Whitin, Analysis of Inventory Systems, Prentice-Hall, Englewood Cliffs, NJ (1963).

[9] Y.L. Kocaga, A. Sen, Spare parts inventory management with demand lead-times and rationing, *IIE Transactions* 39 (2007) 879-898.

[10] P. Melchiors, Restricted time remembering policies for the inventory rationing problem, *International Journal of Production Economics* 81 (2003) 461-468.

[11] P. Melchiors, R. Dekker, M.J. Kleijn, Inventory rationing in an (s, Q) inventory model with lost sales and two demand classes, *Journal of Operational Research Society* 51 (2000) 111-122.

[12] S. Nahmias, W. Demmy, Operating characteristics of an inventory system with rationing, *Management Science* 27 (1981) 1236-1245.

[13] S. Stidham, Stochastic clearing systems, *Stochastic and their applications* 2 (1974) 85-113.

[14] H.T. Teunter, W.K.K. Haneveld, Dynamic inventory rationing strategies for inventory systems with two demand classes, Poisson demand and backordering, *European Journal of Operational Research* 190 (2008)156-178.