Production, Manufacturing and Logistics

A dynamic rationing policy for continuous-review inventory systems

Mehmet Murat Fadıloğlu *, Önder Bulut

Department of Industrial Engineering, Bilkent University, Ankara, Turkey

ARTICLE INFO

ABSTRACT

Stock rationing is an inventory policy that allows differential treatment of customer classes without using separate inventories. In this paper, we propose a dynamic rationing policy for continuous-review inventory systems, which utilizes the information on the status of the outstanding replenishment orders. For both backordering and lost sales environments, we conduct simulation studies to compare the performance of the dynamic policy with the static critical level and the common stock policies and quantify the gain obtained. We propose two new bounds on the optimum dynamic rationing policy that enables us to tell how much of the potential gain the proposed dynamic policy realizes. We discuss the conditions under which stock rationing – both dynamic and static – is beneficial and assess the value of the dynamic policy.

1. Introduction

For inventory systems with distinct customer classes demanding the same item, stock rationing is a well-known tool to differentiate customer classes. Different customers may have different service level requirements or different shortage costs. In such cases, stock rationing allows prioritization of demand classes in order to provide different levels of service and to achieve higher operational efficiency. It is possible to maintain high service levels for certain demand classes while keeping inventory costs at bay by providing lower service levels to certain other demand classes. Demand classes are categorized on the basis of their shortage costs. The highest priority class is the one with the largest unit shortage cost, and the lowest priority class has the smallest unit shortage cost.

Inventory systems subject to multiple demand classes for the same item are frequently observed in real life. Consider a spare parts inventory system. A part can be demanded in order to repair different end products of different importance and criticality. Considering the fact that all demands may not be satisfied instantaneously, demands for spare parts should be prioritized. Moreover, the system may experience urgent orders in case of system breakdowns. The unit shortage cost experience under such a scenario is to be dramatically higher compared to the unit shortage cost of the orders due to the planned maintenance activities. Another example would be a two-echelon inventory system consisting of a warehouse and many retailers. In case of stockout, retailers may place urgent, more critical orders to the warehouse. Furthermore, it may be beneficial to better serve certain retailers that constitute a larger portion of the warehouse's business. In multi-echelon systems, intershipments between the inventory locations in the same echelon may be allowed. However, for any inventory location, direct customer orders have precedence over the intershipment orders that are placed by the other locations.

Customer differentiation is also very important in service sectors. Hotel or airline companies ration their limited capacity according to the priorities of their different customer classes. In this setting, in addition to the rationing decision, another key concern is deciding the prices to be charged to individual customer classes.

For the inventory systems with multiple demand classes, the two common strategies are managing separate individual stocks for each customer class and managing a common stock pool to serve all the classes without any differentiation. The separate stock strategy permits to assign a different service level to each customer class; but it does not allow the system to benefit from risk pooling. The variability of demand is higher in this strategy, thus the system has to hold more safety stock to guarantee the desired service levels. To take the advantage of pooling effect one can use the common stock strategy. However, under the common stock strategy the inventory investment is determined according to the customer class that requires the highest service level. Thus, higher than required service levels are offered to all other customer classes. Inventory rationing is superior to both the separate and the common stock strategies, because as well as using the advantage of the pooling effect, it has the flexibility of providing different service levels to different customer classes.

---

* Corresponding author.
  E-mail address: mmurat@bilkent.edu.tr (M.M. Fadıloğlu).

It is possible to come up with different rationing policies. Yet, the mechanism through which any rationing policy is implemented is to stop serving a lower priority class when the on-hand inventory drops below a certain critical level. The unsatisfied demands are either backordered or lost depending on the nature of the system. Under the critical level only higher priority classes are served and this results in higher service levels for those classes. If there are more than two demand classes, then there has to be more than a single critical level. The critical level for the highest priority class may be assumed to be zero. The critical levels may change dynamically according to the number and the ages of outstanding orders or static threshold levels may be used.

In backordering environments, to completely define the stock rationing policy the way that the backorders are cleared should also be defined. The clearing mechanism specifies how the replenishment orders should be allocated between increasing the stock level and clearing the backorders. The natural way to perform the clearing is to employ the same critical levels for clearing the backorders, i.e., the backorders for a certain customer class are not cleared until the inventory level reaches to the critical level which is associated for that customer class. This mechanism is referred as the priority clearing in the literature. Fadiloglu and Bulut (2005) are the first who analyze the static rationing policy with the priority clearing dynamics.

In the continuous-review setting, there are only a couple of studies that consider a dynamic adjustment of the critical levels for rationing. Since the analysis of rationing systems is complicated even for the static policy, this is a difficult setting. Under the at-most-one-outstanding-order assumption, Teunter and Haneveld (2008) consider a dynamic rationing policy for the backordering environment and Melchiors (2003) considers a so-called time remembering policy for the lost sales case. Except these works, the common practice in the literature is to assume static (time invariant) rationing levels with clearing mechanisms other than the priority clearing (we provide a detailed review of the literature in Section 2). Priority clearing and/or adjusting the critical levels dynamically complicates the analysis considerably.

In this paper, we propose a dynamic rationing policy together with the associated dynamic priority clearing mechanism for continuous-review backordering systems with constant lead-time and unit Poisson demands for two demand classes. The policy uses the age information for all the outstanding orders in order to decide whether a lower priority demand should be satisfied instantaneously or should be backordered (or lost depending on the setting). More specifically, the policy incorporates the outstanding replenishment orders into the on-hand inventory as if they arrive continuously within the lead-time. With the currently available information technologies, it is easy to monitor the status of the outstanding orders and to incorporate the information that they carry into the decision mechanism. Therefore, the proposed policy should not suffer from implementation issues in today's environment.

We conduct a simulation study to evaluate the performance of the proposed policy. Since the analytical evaluation of the policy is not tractable without simplifying assumptions, simulation is the only available tool. In spite of the popularity of simulation in the broader area of operations research and operations management, there is no work in the stock rationing literature that uses simulation to analyze complex, dynamic policies, which outperform static ones. One should also point out that simulation can also be used to estimate any long-run performance measures of the inventory systems under static rationing that are discussed in the literature. Still, authors usually resort to simulation not for direct performance analysis of their models but only for testing their results (e.g., Dekker et al., 1998; Deshpande et al., 2003; Fadiloglu and Bulut, 2005). However, it is possible to obtain the performance measures in any desired confidence interval via simulation, and thus conduct the performance evaluation of the inventory policies under scrutiny.

The rest of the paper is organized as follows. In Section 2, we provide a review of the stock rationing literature. Section 3 discusses the characteristics that the optimal policy should exhibit and proposes two new lower bounds on the optimal policy. In Section 4, we introduce a new class of dynamic rationing policies and discuss the properties a good dynamic rationing policy should exhibit. Based on these properties, we develop a new policy called *Rationing with Exponential Replenishment Flow* (*RERF*). In Section 5, we compare the performance of *RERF* with the static rationing policy and quantify the gain obtained through *RERF* under different scenarios by simulation. The paper concludes in Section 6 with an overall summary and the discussion of the future research directions.

2. Literature review

The research in the area of rationing was initiated in the 1960s. In a periodic-review setting, Veinott (1965) introduces the use of critical levels for providing different service levels to different demand classes. He considers a model with zero lead-time and backordering. Topkis (1968) works on a similar model and proves the optimality of the time remembering critical level policy for both lost sales and backordering cases. His policy is based on dividing each ordering period into a finite number of sub-periods. Via dynamic programming, he finds critical levels for each sub-period. Critical levels are decreasing with the remaining time to the next ordering opportunity. Independent of Topkis (1968), for the lost sales case Evans (1968) and for the backordering case Kaplan (1969) derives the same results for two demand classes. For the multi-period problem, Nahmias and Demmy (1981) consider a (s, S) policy under static rationing with zero lead-time and two demand classes. They derive the expected number of backorders for both customer classes by assuming all demands occur at the end of the period. Moon and Kang (1998) develop a simulation model on the setting of Nahmias and Demmy (1981) with the exception that they assume the arrival process follows a Poisson process. Two recent works that consider periodic-review settings as well are by Frank et al. (2003) and by Atkins and Katircioglu (1995).

Nahmias and Demmy (1981) are also the first to consider stock rationing in continuous-review environment. They assume two demand classes with unit Poisson arrivals, constant lead-time and full backordering for performance evaluation purposes. They derive approximate expressions for the expected number of backorders and for the fill rates for both classes under a (Q, r, K) policy, which is a (Q, r) policy with the fixed critical level K. Their approximation is based on the at-most-one-order-outstanding assumption. Moon and Kang (1998) consider compound Poisson demand arrivals and provide a simulation study on the setting of Nahmias and Demmy (1981).

Dekker et al. (1998) consider the same setting with Nahmias and Demmy (1981) with $Q = 1$. Without specifying any clearing mechanism, they derive the exact fill rate expression for the non-critical demand class and make an approximation for the critical class fill rate by conditioning on the time that stock level hits the critical level. They test their approximation under three different clearing mechanisms using simulation. Kocaga and Sen (2007) extend the approximation of Dekker et al. (1998) to accommodate a demand lead-time for non-critical orders. They conduct a simulation study to assess the performance of the approximation.

Deshpande et al. (2003) consider exactly the same setting with Nahmias and Demmy (1981). They derive approximate service level expressions that are exact under a threshold clearing mechanism they construct to make the analysis possible. It is interesting that their analysis yields the same results with Dekker et al. (1998) when $Q = 1$. They also provide an algorithm to find the optimal policy parameters under a given cost structure. They simulate the rationing policy with the priority clearing mechanism to compare the performance of the threshold clearing with the performance of priority clearing. Zhao et al. (2005) analyze a game theoretical model of a decentralized dealer network in which each dealer can share its inventory with the others. They use the threshold clearing mechanism that Deshpande et al. (2003) introduce.

Arslan et al. (2005) analyze the multiple demand classes extension of the setting that Nahmias and Demmy (1981) consider. They construct an equivalent multi stage serial system of the original single location system. For the ease of analysis, they assume that in each stage the backorders are cleared in the order of occurrence. Thereby, they derive approximate results for the priority clearing mechanism. Fadil-oglu and Bulut (2005) also consider the same setting. They propose a method which captures the priority clearing dynamics for continuous-review inventory systems with backordering under static rationing policy. They assume two demand classes with Poisson arrivals and constant lead-time. They sample the continuous system at multiples of the lead-time and show that the state of the system evolves according to an embedded Markov chain. They provide a recursive procedure to obtain the transition probabilities of the embedded chain and obtain the steady-state probabilities of interest with desired accuracy by considering a truncated version of the chain.

Dekker et al. (2002) consider a lot-per-lot continuous-review setting with the static critical level rationing in a lost sales environment. The clearing mechanism is not relevant in the lost sales case. They provide exact expressions for service levels under general stochastic lead-time and multiple demand classes. Their results are adapted from the analysis of $M/G/\infty$ queue under state-dependent arrival rates. Melchiors et al. (2000) also analyze a lost sales environment under (Q, r) policy with the at-most-one-order-outstanding assumption. This analysis is exact when $r < Q$. Melchiors (2003) extends this model by considering a time remembering rationing policy that allows different critical levels for different time slots between the placement of an order and its arrival. Critical levels are set according to the age of the outstanding order. This is a restricted dynamic policy because critical levels are assumed to be constant over predetermined time intervals. However, the unknown optimal policy should allow the critical levels change at any point in time. Teunter and Haneveld (2008) also consider a time remembering policy for the backordering case. They aim to determine the set of critical remaining lead-time values $(L_1, L_2 \ldots)$ for the rationing decision. If the remaining lead-time is less than L_1 they do not ration the stock, if it is between L_1 and $(L_1 + L_2)$ one item is reserved for the high priority class and so on. Under the at-most-one-outstanding-order assumption, they approximate the optimal critical remaining lead-time values. Using two examples, they demonstrate that the dynamic policy outperforms the static policy.

Stock rationing problems are also analyzed in continuous-time systems with capacitated replenishment channel. Ha (1997a) considers a make-to-stock production facility with a single exponential server, zero setup cost, multiple demand classes and lost sales. He shows that the lot-for-lot policy is optimal for production and the static critical level policy is optimal for stock rationing. Ha (2000) extends the work of Ha (1997a) to Erlangian production times. He proves that a critical work storage level policy is optimal where work storage level is the number of completed Erlang stages for the items in the system. Ha (1997b) analyzes the same setting with Ha (1997a) with the exception of considering a backordering environment with two demand classes. He characterizes the optimal rationing policy as a critical level policy where the level is decreasing in the number of backorders of the non-critical class. Vericourt et al. (2002) extends the findings of Ha (1997b) to multiple demand classes. Gayon et al. (2006) considers Erlang production times in a backordering environment.

3. Dynamic rationing

In the stock rationing literature, it is well documented that the performance of the static rationing policy can be improved by utilizing the information on the status of the outstanding replenishment orders. For a given set of policy parameters, i.e., given (Q, r, K) values, although the static policy would not allow it, one should prefer to satisfy an arriving lower priority customer instantaneously if an outstanding order is about to arrive and increase the inventory level. A similar scenario can also be entertained for the backorder clearing mechanism of the rationing policy; it may be profitable to clear some backorders of lower priority class before the inventory level is increased above the threshold rationing level K. Therefore, it should be clear that the optimal rationing policy should be a dynamic policy that allows the threshold rationing level to change in time depending on the number and ages of outstanding orders. However, the characterization of this optimal policy structure would be quite hard, if possible at all. Such a policy would have to depend on the time-to-arrive of all outstanding orders, which is a random variable. Thereby, it is very hard to analyze, if possible, any dynamic policy without simplifying assumptions.

Teunter and Haneveld (2008) consider dynamic stock rationing but the analysis is based on the assumption of at-most-one-outstanding-order and is computationally tractable only for limited settings. The difficulty in the analysis of dynamic policies mainly arises from the fact that one should incorporate the ages (or the time-to-arrive values) of all outstanding orders into the system state definition. Moreover, since the number of outstanding orders changes in time depending on the realizations of the demand processes and the policy parameters, the size of the state vector itself is a random variable. Thus, we evaluate the performance of the policy we propose, which called *Rationing with Exponential Replenishment Flow* (*RERF*) via simulation.

Although the structure of the optimal rationing policy is unknown, we would like to tell something about the quality of the policy we suggest with respect to the optimal policy. Since it is not possible to do the performance evaluation of an unknown policy, we develop two lower bounds for any possible rationing policy (static or dynamic). Using these bounds, we are able to assess how much our policy realizes out of the maximum possible potential for the rationing policies. Moreover, these bounds also point out the settings at which there is no meaningful gain to be obtained by applying any kind of dynamic policy.

Consider the inventory system discussed in Section 1 with parameters $\lambda_i, \pi_i, \hat{\pi}_i, i \in \{1, 2\}$, and h, A, i.e., the arrival rates, the unit backordering costs, the time-dependent backordering costs, the holding cost rate and the fixed ordering cost. Without loss of generality we assume $\pi_1 > \pi_2$ and $\hat{\pi}_1 > \hat{\pi}_2$, which means class 1 is the higher priority demand class. We denote this system with O. Based on this system; we can construct two related inventory systems, each of which is subject to a single-demand class in order to obtain lower bounds. The first system we propose is the one with demand rate $\lambda_1 + \lambda_2$, and backorder costs $\pi_2, \hat{\pi}_2$. The second one is another single-demand class system with demand rate λ_1 and backorder costs $\pi_1, \hat{\pi}_1$. We denote the first system with N_1, and the second one with N_2. For each of these new

systems we assume that the holding cost rate and the fixed ordering cost are the same with the original two demand class system. Since N_1 is constructed using the minimum of the backorder costs of the two demand classes (while keeping the total demand rate constant), the optimal long-run average cost of N_1 is a lower bound on the long-run average cost of O using an optimal dynamic rationing policy. N_2 constitutes another lower bound, since class 2 is completely disregarded and it is not possible for any rationing policy to operate without experiencing any cost related to class 2. Note that for both systems, since there is only a single-demand class, there is no need for any kind of rationing. To obtain the above mentioned lower bounds, one only needs to optimize N_1 and N_2 with respect to the parameters Q and r.

Any one of the two bounds may be tighter than the other for a given problem instance. Therefore, the strategy is computing both of the bounds and taking the maximum as the lower bound on the cost of the optimal policy. However, it is more likely that N_1 provides a tighter bound when a high proportion of the total demand is due to class 2. Note that for a given total demand rate $\lambda = \lambda_1 + \lambda_2$, the lower bound provided by N_1 is independent of the demand mix, i.e., it is the same for all $p_1 = \lambda_1/\lambda$ values. On the other hand, when $p_1 = 1$ the lower bound provided by N_2 takes its maximum value (which is certainly greater than the lower bound obtained by N_1, because when $p_1 = 1$ both of the systems experience the same demand rate but N_2 assumes higher backorder costs) and it decreases down to zero as p_1 goes to zero. Therefore, depending on the values of the total demand rate and the cost parameters, the lower bounds obtained by the systems N_1 and N_2 should intersect at a p_1 value in $[0,1]$. Thus, for the values of p_1 between zero and the intersection point the lower bound provided by N_1 should be tighter and for the other possible values of p_1 (from the intersection point to 1) the lower bound obtained by N_2 should be tighter.

At this point, it should also be noted that Deshpande et al. (2003) provides a different lower bound on the cost of any rationing policy. Their bound is based on an approximate analysis for the priority clearing mechanism. However, the simulation study considered in Section 5 illustrates that our bound is tighter than the one suggested by Deshpande et al. (2003).

4. Rationing with continuous replenishment flow

The current level of sophistication in information and computer technologies enables us to consider more elaborate policies. Although the analyses of these elaborate policies are difficult and mostly intractable, it is still possible to estimate the steady-state behavior of the system with simulation in a reasonable amount of time with the current computer speeds. Hence, we propose a dynamic rationing policy that makes use of all the available system information at any point in time and evaluate the performance of the policy by a simulation study.

In this section, we introduce our policy with its dynamic priority clearing mechanism in a continuously reviewed inventory setting where the replenishment orders are placed according to the (Q,r) policy. The (Q,r) policy dictates that a batch of Q units is ordered whenever inventory position hits the reorder level r. Inventory position is the sum of inventory level (on-hand inventory minus the number of backorders) and ordered units that are in the replenishment channel. This means that the inventory position increases at ordering points, whereas inventory level increases when the orders arrive. We assume two customer classes and a deterministic supply lead-time, $L > 0$. Demand arrivals are generated by two independent stationary Poisson processes with rates λ_1 and λ_2, respectively, for class 1 and class 2 customers.

Static rationing policies, which are easier to analyze and implement, only utilize part of the information on system state. In the static policies, the replenishment and rationing decisions are based on the inventory position and the on-hand inventory level, respectively. However, unless the replenishment lead-time is memoryless, i.e., exponential, there is additional information available about the arrival times of the orders in the replenishment channel. In the case of deterministic lead-time we consider, the arrival times are exactly known once the orders are placed. Thus, a "good" rationing policy should exploit this information to extract value.

Even if the on-hand inventory level is below the (static) critical level, it is better to instantaneously satisfy a class 2 demand if the outstanding replenishment orders are about to increase the on-hand inventory level. If a replenishment arrival will happen after a short duration, the likelihood of a class 1 demand arrival before the replenishment is relatively small. Note that the higher is the class 1 arrival rate, the shorter the duration has to be. Thus, the critical level should be lowered as the ages of the replenishment orders increase.

Instead of defining the critical level as a function of the ages of outstanding orders, we opt to use a constant critical level K and a modified on-hand stock level, which are adjusted dynamically utilizing the information on the outstanding replenishment orders. Note that although the two approaches are basically equivalent, the latter better lends itself to interpretation.

We define a new class of policies called *Rationing with Continuous Replenishment Flow* (*RCRF*) that uses a constant critical rationing level on the modified on-hand inventory, which incorporates the outstanding orders to the on-hand inventory as if they are arriving continuously within the lead-time. The only difference between *RCRF* and the static rationing policy is the variable on which the rationing mechanism is defined.

Let $X(t)$ denote the number of outstanding replenishment orders at time t, $a_i(t)$ denote the age of ith oldest outstanding replenishment order at time t where $0 \leqslant a_i(t) \leqslant L, 1 \leqslant i \leqslant X(t)$, and $OH(t)$ denote the on-hand inventory level at time t. We define the modified on-hand inventory level at time t, $OH_m(t)$, as

$$OH_m(t) = \begin{cases} OH(t) + Q \sum_{i=1}^{X(t)} f(a_i(t)), & OH(t) > 0, \\ 0, & OH(t) = 0. \end{cases} \tag{1}$$

We do not modify the on-hand inventory when it is zero, since any incoming demand has to be backordered. Different *RCRF* policies correspond to different families of $f(\cdot)$ functions, satisfying the following properties:

1. $f(t) : [0, L] \rightarrow [0, 1]$,
2. $f(t)$ is an increasing function of t,
3. $f(t) = 1$ for $t = L$.

(1) implies that the contribution of ith outstanding order to the modified on-hand inventory level at time t is $Qf(a_i(t))$. Thus, $f(a_i(t))$ should be considered as the fraction of the replenishment order quantity that is incorporated to the on-hand inventory when the age of the outstanding order is $a_i(t)$. Obviously, as the age of the outstanding order increases this fraction should increase up-to 1 (it should be 1 when the outstanding order actually arrives, i.e., when $a_i(t) = L$).

According to *RCRF*, if there is no outstanding order at time t, we have $OH_m(t) = OH(t)$. In this circumstance, we compare the on hand stock level with the critical level to make the rationing decision as in the static critical level policy. If $OH(t) > K$, then $OH_m(t)$ is certainly greater than K. Thus an arriving class 2 demand is satisfied. If $0 < OH(t) \leqslant K$, the arriving class 2 demand is satisfied instantaneously provided that $OH_m(t) > K$. For the other cases, class 2 demand is backordered. If the arriving demand belongs to class 1, it is satisfied instantaneously if $OH(t) > 0$ as in the case of the static policy. Otherwise, it is backordered.

In order to complete the definition of *RCRF*, the modified priority clearing mechanism should also be specified. Suppose a replenishment order arrives at time t. After clearing class 1 backorders (if there is any), the remaining replenishment order quantity (if any remains) first used to increase the modified on-hand level up-to K and then class 2 backorders are cleared.

Any flow function $f(t)$ that satisfies the stated three properties can be used to modify the on-hand inventory according to (1). However, at this stage, we take a further step and propose to consider only the *RCRF* policies that assume strictly convex flow function. Because, the value of the information gained from the outstanding orders should diminish as we go from the oldest to the youngest order. In other words, the change in the impact of the orders on modified on-hand as we move in time, should be greater for those orders that are closer to arrive. Consequently, the difference between $OH_m(t)$ and $OH(t)$ should be mostly due to the outstanding orders that are to arrive in the very near future.

The above reasoning can be clarified with an example. Suppose that the flow function has the form $f(a_i(t)) = \left(\frac{a_i(t)}{L}\right)^n$ and at time t' there are two outstanding orders, $X(t') = 2$. Under these conditions, consider the following two cases: In the first case, $\frac{a_1(t')}{L} = 0.9$, $\frac{a_2(t')}{L} = 0.1$ and in the second one $\frac{a_1(t')}{L} = 0.6$, $\frac{a_2(t')}{L} = 0.4$. When $n = 1$, (1) returns the same $OH_m(t')$ values for both cases, because $\sum_{i=1}^{X(t)} f(a_i(t')) = 1$ for both of them. However, when $n = 2$, i.e., when the flow function is strictly convex, $\sum_{i=1}^{X(t)} f(a_i(t'))$ is 0.82 for the first case and it is 0.52 for the second one. Thus, $OH_m(t')$ is larger for the first case and the decision maker is more eager to satisfy a class 2 demand that arrives at t'. The unknown optimal policy would also distinguish the two cases and would be less conservative in satisfying an arriving class 2 demand for the first case than for the second case. Because, in the first case, there is an outstanding order that is close to arrive (compared to the other outstanding orders considered in the example), it is very likely that class 1 backorders in the near future can be avoided.

Another important criterion in choosing an appropriate policy from the *RCRF* class is the value of the flow function for an outstanding order that is just placed. For the flow function discussed in the above example, we have $\lim_{a_i(t)\to 0^+} f(a_i(t)) = 0$. However, the flow function should assume a positive value right at the onset, because the information that there is an outstanding order should have a nonzero value. There may be situations where we would choose to satisfy an arriving class 2 demand if we know a replenishment order has just been placed and will arrive after a lead-time period (for example if the lead-times are extremely small).

In the light of the above discussion about the intuitive criteria, we propose (2) as the flow function and define the policy *Rationing with Exponential Replenishment Flow (RERF)*. As a member of *RCRF* class *RERF* modifies the on-hand inventory according to (1).

$$f(a_i(t)) = e^{-\lambda_1(L-a_i(t))n}. \tag{2}$$

In (2), $(L - a_i(t))$ is the remaining time to arrive for the ith oldest outstanding order and $e^{-\lambda_1(L-a_i(t))}$ is the probability that there will be no class 1 demand arrivals until the arrival of the ith oldest outstanding order. Since (2) is a decreasing function of λ_1, i.e., the probability of zero class
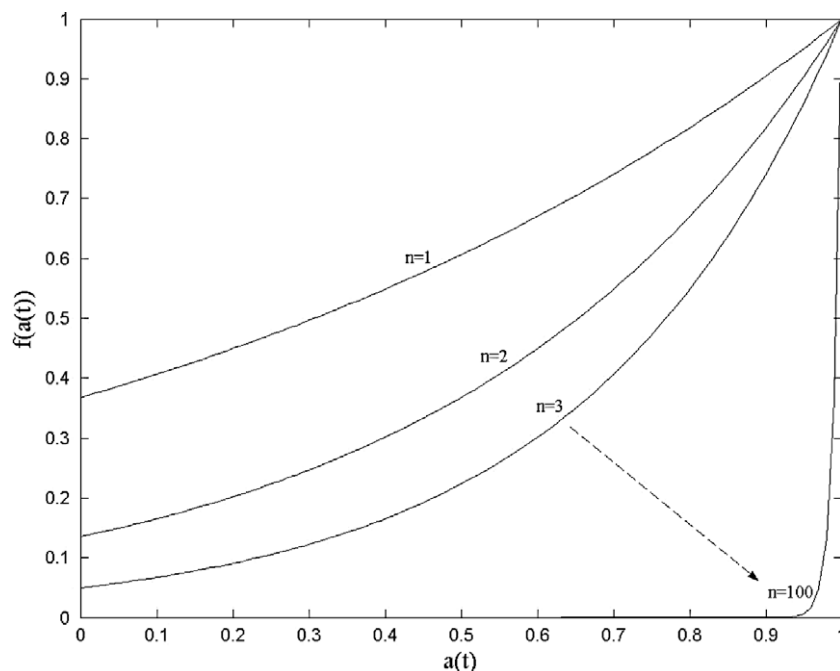

Fig. 1. Exponential flow function ($L = 1, \lambda = 1$).

1 demand arrival in a specific length of time decreases as λ_1 increases, if class 1 arrival rate is high the modified on-hand inventory is close to the real on-hand level and not many extra class 2 demands are satisfied compared to the static rationing. On the contrary, if λ_1 is low then much more class 2 demands can be satisfied. That is to say, *RERF* updates the on-hand inventory by considering the risk of having class 1 backorders in the future.

Power n is a constant that is used to fine-tune the policy. It is a parameter to be optimized together with Q, r and K. As n increases the value of the information gained from the outstanding orders diminishes and the diminishing rate is faster for the younger orders. Fig. 1 illustrates this situation for $L = 1, \lambda = 1$. As $n \to \infty$, *RERF* does not utilize any information that the outstanding orders carry. Thus, it becomes identical with the static critical level policy, i.e., $\lim_{n\to\infty} OH_m(t) = OH(t)$ for all t. This observation guarantees that *RERF* would perform at least as good as the static policy. For the same system parameters, the optimal long-run average cost of the static policy should be an upper bound for the optimal long-run average cost of *RERF*.

It should be also noted that, for the same (Q, r, K) values, *RERF* increases class 2 fill rate and decreases class 1 fill rate compared to the static policy. Even though the real on-hand stock level is at or below K, *RERF* satisfies class 2 customers if the modified on-hand level is above K. Thus, the reserve stock allocated for class 1 demands decreases.

In this study, we only consider the case of two customer classes. However, it is straightforward to adapt *RERF* to the case with n customer classes by defining a different flow function for each class with the exception of class 1 (class 1 demands are always satisfied whenever there is stock on-hand). For $j \in \{2, 3, \ldots, n\}$, (3) defines the flow function for class j that consider the risk of having backorders from the higher priority customer classes within the remaining lead-time of ith outstanding order

$$f_j(a_i(t)) = e^{-\left(\sum_{k=1}^{j-1} \lambda_k\right)(L-a_i(t))n}. \tag{3}$$

This function is employed in computing the modified on-hand inventory level (see (1)) to be used for the rationing decision at the arrival time of jth class customer. There are $n - 1$ critical levels for classes other than class 1. The modified inventory level is compared with the critical level for class j to decide whether to ration or not.

5. Performance evaluation of RERF via simulation

Simulation is one of the best available tools for the analysis of complex systems for which analytical techniques are not tractable. It enables us to compute any performance measure of interest for systems under the policy of consideration. For any parameter setting, it is possible to find the optimal policy parameters using simulation–optimization. Hence, to evaluate the performance of the policy proposed in the previous section, we developed a simulation model and programmed in C. We identified the optimum policy parameters and quantified the gain obtained through *RERF* under different scenarios by comparing the static critical level policy. To distinguish the cases where stock rationing is beneficial and to assess the relative value of *RERF*, we also simulated the common stock policy, which provides an upper bound on the costs of the rationing policies, and compared its performance with the static policy. Moreover, to characterize the performance of *RERF* relative to all possible rationing policies and to identify the conditions under which dynamic stock rationing is valuable, for some problem instances we simulated the single-class systems N_1 and N_2 (described in Section 3) and obtained two lower bounds on the performance of the unknown optimal rationing policy.

For each problem instance, we identified the policy parameters ((Q, r, K, n) for *RERF*, (Q, r, K) for the static policy and (Q, r) for the common stock policy) that provide the minimum long-run average cost estimate. We run the simulation model of each inventory system for 600,000 customer arrivals to ensure the stability of the estimates. To verify our results, we compared the exact class 2 fill rate attained under the static policy, which is provided by Deshpande et al. (2003), with the simulation estimate for each instance of our problem set. The largest deviation of simulation estimate from the exact value is 0.0044, i.e., class 2 fill rate estimate is exact to the second significant digit. Therefore, we concluded that for each problem instance "600,000 customer arrivals" provide a reasonable run-length to observe the steady-state conditions of the inventory systems named above.

For given policy parameters, the long-run average cost expression of any policy can be written as

$$C = A\frac{\lambda_1 + \lambda_2}{Q} + h\bar{I} + \pi_1(1 - \beta_1)\lambda_1 + \pi_2(1 - \beta_2)\lambda_2 + \hat{\pi}_1\gamma_1\lambda_1 + \hat{\pi}_2\gamma_2\lambda_2, \tag{4}$$

where A is the fixed ordering cost, h is the holding cost rate, π_i is the unit backorder cost and $\hat{\pi}_i$ is the time-dependent backordering cost. In addition, $\bar{I}$ denotes the average inventory, β_i denotes the fill rate for class i and γ_i denotes the average backorder time per customer for class $i (i = 1, 2)$. Let $\tilde{C}_{RERF}, \tilde{C}_{SP}$ and $\tilde{C}_{CS}$ denote the long-run average cost estimate of *RERF*, the static policy and the common stock policy correspondingly. Then, the performance gain of *RERF* as the percent cost reduction obtained by operating the system under *RERF* instead of the static rationing policy with the optimal policy parameters can be defined as

$$G_{RCRF} = 100 \frac{\left(\begin{array}{l} \min_{(Q,r,K)}\{\tilde{C}_{SP} : r + Q > K \geqslant 0, Q \geqslant 1, r \geqslant 0, \text{all integer}\} \\ -\min_{(Q,r,K,n)}\{\tilde{C}_{RCRF} : r + Q > K \geqslant 0, Q \geqslant 1, r \geqslant 0, n \geqslant 1, \text{all integer}\} \end{array}\right)}{\min_{(Q,r,K)}\{\tilde{C}_{SP} : r + Q > K \geqslant 0, Q \geqslant 1, r \geqslant 0, \text{all integer}\}}. \tag{5}$$

Similarly, the performance gain of the static rationing policy as the percent cost reduction relative to the common stock policy is

$$G_{SP} = 100 \frac{\left(\begin{array}{l} \min_{(Q,r)}\{\tilde{C}_{CS} : Q \geqslant 1, r \geqslant 0, \text{all integer}\} \\ -\min_{(Q,r,K)}\{\tilde{C}_{SP} : r + Q > K \geqslant 0, Q \geqslant 1, r \geqslant 0, \text{all integer}\} \end{array}\right)}{\min_{(Q,r)}\{\tilde{C}_{CS} : Q \geqslant 1, r \geqslant 0, \text{all integer}\}}. \tag{6}$$

To obtain G_{RCRF} and G_{SP} values for the settings that reflect the main trade-offs we generated 288 problem instances varying in total demand rate, ratio of class 1 demand rate to the total rate, unit and time-dependent backorder costs of both classes and the setup cost. For the lost

sales case, we used the same problem instances excluding the time-dependent shortage costs. Since we considered different values of the other cost parameters, we fixed the unit holding cost to 5 without loss of generality. We also fixed the lead-time to the unit-time without loss of generality. The problem instances considered in the simulation study were formed from the elements of the following sets:

$$\lambda = (\lambda_1 + \lambda_2) \in \{25, 5\}, \quad p_1 = \frac{\lambda_1}{\lambda} \in \{0.1, 0.5, 0.9\}, \quad \pi_1 \in \{2, 10\}, \quad r_1 = \frac{\pi_1}{\pi_2} \in \{5, 1.25\}, \quad \hat{\pi}_1 \in \{1, 5\},$$

$$\hat{r}_1 = \frac{\hat{\pi}_1}{\hat{\pi}_2} \in \{5, 1.25\}, \quad A \in \{0, 2, 10\}, \quad L \in \{1\}, h \in \{5\}.$$

5.1. Backordering case

Tables 1–3 compare the performance of the policies for the backordering case. The data on each row of Table 1 (the maximum and the average percent gain) were obtained using 16 different settings varying in $(\pi_1, \pi_2, \hat{\pi}_1, \hat{\pi}_2)$ values. As seen from Table 1, for any given total demand rate and the setup cost, the benefit of the static rationing over the common stock policy is maximized when the total demand rate is evenly distributed among the classes, i.e., $\lambda_1 = \lambda_2$. This is an expected result, because as p_1 goes to 0 or 1, the value of the static rationing should diminish. As p_1 decreases, i.e., as the portion of class 2 demand gets higher and higher, the static policy lowers the threshold rationing level towards 0 in order to prevent large numbers of class 2 backorders. When $p_1 = 0.1$, almost for all the cases the optimal rationing level is 0. Therefore, cost reduction is obtained only through the priority clearing mechanism, which decreases γ_1 and increases γ_2 compared to the first-come-first-served basis clearing. On the other extreme, as p_1 goes to 1, K is increased and most of the stock is reserved for the class 1 customers. When $p_1 = 0$ or $p_1 = 1$, the static rationing policy, RERF and the common stock policy all become identical, because in these cases there is only one customer class in the system and any kind of class differentiation is irrelevant.

Table 1
Comparison of policies for all possible shortage cost pairs.

			G_{RERF}		G_{SP}	
			Max	Avg.	Max	Avg.
$\lambda = 25$	$A = 0$	$p_1 = 0.10$	10.10	5.43	4.49	1.26
		0.50	8.95	4.63	29.76	12.01
		0.90	1.35	0.79	12.45	6.39
	$A = 2$	$p_1 = 0.10$	3.67	1.46	2.90	0.58
		0.50	8.73	3.06	13.51	4.71
		0.90	1.13	0.55	6.84	2.99
	$A = 10$	$p_1 = 0.10$	1.49	0.52	3.47	0.83
		0.50	2.84	0.84	7.40	2.65
		0.90	0.46	0.19	3.27	1.29
$\lambda = 5$	$A = 0$	$p_1 = 0.10$	2.21	0.51	6.88	1.18
		0.50	2.61	0.94	12.48	4.42
		0.90	0.64	0.23	5.55	2.34
	$A = 2$	$p_1 = 0.10$	0.51	0.25	1.72	0.28
		0.50	2.75	0.83	8.49	2.38
		0.90	0.88	0.32	3.60	1.41
	$A = 10$	$p_1 = 0.10$	0.21	0.18	1.98	0.53
		0.50	0.87	0.32	4.32	1.56
		0.90	0.17	0.09	2.02	0.89

Table 2
Comparison of policies for $\lambda = 25$, $A = 0$.

		$p_1 = 0.10$		$p_1 = 0.50$		$p_1 = 0.90$	
		G_{RERF}	G_{SP}	G_{RERF}	G_{SP}	G_{RERF}	G_{SP}
$\pi_1 = 10, \hat{\pi}_1 = 5$	$r_1 = 5, \hat{r}_1 = 5$	8.81	2.53	7.79	18.60	0.76	8.08
	$r_1 = 5, \hat{r}_1 = 1.25$	7.18	0.36	3.58	14.87	1.35	6.87
	$r_1 = 1.25, \hat{r}_1 = 5$	3.23	0.09	3.96	1.69	1.00	1.97
	$r_1 = 1.25, \hat{r}_1 = 1.25$	3.12	0.00	2.87	2.06	0.79	1.06
$\pi_1 = 10, \hat{\pi}_1 = 1$	$r_1 = 5, \hat{r}_1 = 5$	10.10	4.10	8.95	19.31	0.61	9.23
	$r_1 = 5, \hat{r}_1 = 1.25$	8.65	3.70	8.49	17.86	0.40	8.92
	$r_1 = 1.25, \hat{r}_1 = 5$	2.87	0.00	3.91	1.96	1.00	2.28
	$r_1 = 1.25, \hat{r}_1 = 1.25$	2.47	0.51	3.17	3.02	0.33	2.77
$\pi_1 = 2, \hat{\pi}_1 = 5$	$r_1 = 5, \hat{r}_1 = 5$	5.12	4.49	3.60	25.02	0.63	10.96
	$r_1 = 5, \hat{r}_1 = 1.25$	0.82	0.79	0.42	9.87	1.14	3.90
	$r_1 = 1.25, \hat{r}_1 = 5$	5.80	0.60	5.97	8.83	0.79	6.63
	$r_1 = 1.25, \hat{r}_1 = 1.25$	2.19	0.52	2.40	2.94	0.97	2.37
$\pi_1 = 2, \hat{\pi}_1 = 1$	$r_1 = 5, \hat{r}_1 = 5$	9.86	1.92	3.67	29.76	1.03	12.45
	$r_1 = 5, \hat{r}_1 = 1.25$	4.42	0.00	3.18	21.22	0.32	10.17
	$r_1 = 1.25, \hat{r}_1 = 5$	7.15	0.00	7.11	8.29	0.74	8.00
	$r_1 = 1.25, \hat{r}_1 = 1.25$	5.10	0.57	5.02	6.83	0.85	6.57

As expected, *RERF* outperforms the static rationing policy at each parameter setting. Parallel to the above discussion about the static rationing, for given λ and A, the benefit of *RERF* over the static rationing policy appears to be maximized when p_1 is far away from the extreme cases and close to 0.5. The only exception is observed when $\lambda = 25$ and $A = 0$. In this case, *RERF* provides the most significant cost reduction, 10.10% in the best case and 5.43% on the average, when the large proportion of the demand is from class 2. This shows that although the static rationing seems to be a valuable policy when the demand rates of the classes are close to each other, rationing the stock dynamically can provide substantial extra benefit for such cases (especially when the class 1 backorder costs are much more higher than the class 2 backorder costs). In addition to that, dynamic rationing is also a valuable tool when λ_2 is high compared to λ_1. In general, when p_1 is lower, the static policy does not perform well as discussed in the previous paragraph. However, for such cases *RERF* provides considerable additional savings especially when total demand rate is high. This is due to the fact that *RERF* has the capability of increasing β_2 and decreasing γ_2, i.e., decreasing β_1 and increasing γ_1, values that attained under the static policy. Naturally, this capability provides significant gain when λ_2 is high compared to λ_1.

On the other hand, when p_1 is high, *RERF* does not provide noteworthy additional savings. This can be explained with the fact that when the demand is mostly generated by class 1, keeping γ_1 and β_1 values at the levels dictated by static policy is much more beneficial than decreasing them by satisfying some class 2 demands earlier. Hence, assuming continuous replenishment flow and modifying the inventory level accordingly (to satisfy some class 2 demands earlier) is not useful when p_1 is high. In such settings, the information that outstanding orders provide does not have any significance. However, this is true not only for *RERF* but also for all possible dynamic policies as demonstrated by the bounds in Fig. 2.

Fig. 2 summarizes all the above discussions related to Table 1 for the case $\lambda = 25, \pi_1 = 10, \pi_2 = 2, \hat{\pi}_1 = 1, \hat{\pi}_2 = 0.2, h = 5, A = 2$. It clarifies the impact of p_1 on the performance of *RERF* and the static policy. In the figure, in addition to the optimal costs of common stock, static and *RERF* policies, the lower bounds obtained from the systems N_1, N_2 are exhibited. Lower Bound 1 and Lower Bound 2 are the optimal long-run average costs of systems N_1 and N_2, respectively. For each case, the lower bound we propose should be considered as the maximum of Lower Bound 1 and Lower Bound 2. It can be seen from Fig. 2 that when $p_1 \geqslant 0.7$ none of the dynamic rationing policies can provide any extra gain over the static policy, which explains why *RERF* could not provide meaningful savings when $p_1 = 0.9$. For the other cases, i.e., $p_1 < 0.7$, the extra benefit of *RERF* (the benefit of *RERF* over the static policy) is almost same as the benefit of the static policy over the

Table 3
Comparison of policies for $\lambda = 25, A = 2$.

|		$p_1 = 0.10$		$p_1 = 0.50$		$p_1 = 0.90$	
		G_{RERF}	G_{SP}	G_{RERF}	G_{SP}	G_{RERF}	G_{SP}
$\pi_1 = 10, \hat{\pi}_1 = 5$	$r_1 = 5, \hat{r}_1 = 5$	3.40	1.69	7.06	9.01	1.04	4.57
	$r_1 = 5, \hat{r}_1 = 1.25$	3.13	0.46	3.98	7.05	0.95	3.51
	$r_1 = 1.25, \hat{r}_1 = 5$	1.21	0.21	2.35	0.00	0.94	0.53
	$r_1 = 1.25, \hat{r}_1 = 1.25$	1.23	0.00	2.01	0.00	1.08	0.00
$\pi_1 = 10, \hat{\pi}_1 = 1$	$r_1 = 5, \hat{r}_1 = 5$	3.45	2.90	8.73	9.52	0.53	5.84
	$r_1 = 5, \hat{r}_1 = 1.25$	3.67	1.62	7.60	8.49	0.57	5.03
	$r_1 = 1.25, \hat{r}_1 = 5$	0.98	0.00	2.10	0.37	0.56	0.24
	$r_1 = 1.25, \hat{r}_1 = 1.25$	1.38	0.00	1.95	0.09	1.13	0.00
$\pi_1 = 2, \hat{\pi}_1 = 5$	$r_1 = 5, \hat{r}_1 = 5$	1.92	1.33	1.29	12.53	0.20	5.91
	$r_1 = 5, \hat{r}_1 = 1.25$	0.39	0.00	0.96	3.02	0.00	2.20
	$r_1 = 1.25, \hat{r}_1 = 5$	0.38	0.35	3.30	1.95	0.42	2.86
	$r_1 = 1.25, \hat{r}_1 = 1.25$	0.18	0.00	0.80	0.33	0.37	0.77
$\pi_1 = 2, \hat{\pi}_1 = 1$	$r_1 = 5, \hat{r}_1 = 5$	0.90	0.88	0.57	13.51	0.17	6.84
	$r_1 = 5, \hat{r}_1 = 1.25$	0.41	0.00	1.05	8.38	0.04	4.60
	$r_1 = 1.25, \hat{r}_1 = 5$	0.24	0.00	3.49	0.94	0.44	2.84
	$r_1 = 1.25, \hat{r}_1 = 1.25$	0.43	0.00	1.78	0.19	0.27	2.06
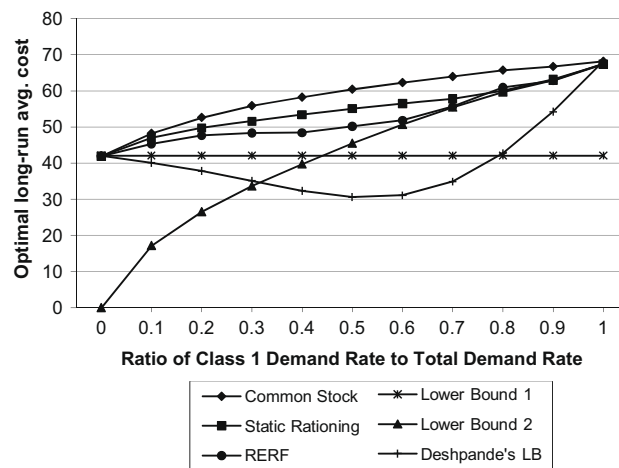
Fig. 2. Impact of demand mix.

common stock policy. The lower bound proposed by Deshpande et al. (2003) is also provided in Fig. 2. It is apparent that the lower bound proposed in this paper is tighter than the one proposed by Deshpande et al. (2003).

Table 1 also shows that as the total traffic rate increases both the static rationing and *RERF* provides considerable cost savings. Especially, when p_1 is 0.1 or 0.5 and the setup cost is low, i.e., A is 0 or 2, and $\lambda = 25$, *RERF* results in remarkable additional savings. Increasing the total traffic rate sharpens the trade-off between holding and shortage costs. Since the value of rationing is based on this trade-off, it increases with the traffic rate. Moreover, we observe that as λ increases, $\lambda L/Q$, which is the expected number of outstanding orders, increases in most cases. Since *RERF* assumes continuous flow of the outstanding replenishment orders, it has more capability to re-optimize the parameters and to increase the cost saving as the number of outstanding orders increases. Fig. 3 generalizes this discussion by considering ten different total demand rate values for the case $p_1 = 0.5, \pi_1 = 10, \pi_2 = 2, \hat{\pi}_1 = 1, \hat{\pi}_2 = 0.2, h = 5, A = 2$. As λ increases, both the benefit of static policy over the common stock policy and the benefit of *RERF* over the static policy increase. Moreover, the gap between *RERF* and Lower Bound 2 is not very large, which means that our policy realizes most of the existing potential for the rationing policies. Here it should be also noted that the actual gap between *RERF* and the unknown optimal policy should be much less than the gap between *RERF* and the lower bounds. If we consider the case where $p_1 = 0.9$, the benefit of unknown optimal rationing policy over the static policy is almost zero for all λ values as shown in Fig. 4, i.e., the curves corresponding to the static policy and Lower Bound 2 overlap.

Setup cost is the other important cost parameter that affects the performance of the policies. In an environment with less setup cost the effect of the service level differentiation is more important. As the setup cost increases, we observe that the average ordering cost appears to dominate the average holding and shortage costs. Since the rationing policies derive benefit from the trade-off between the average
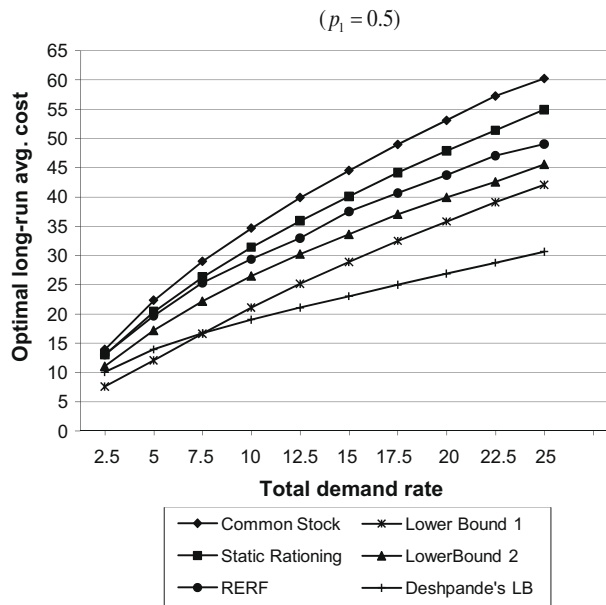

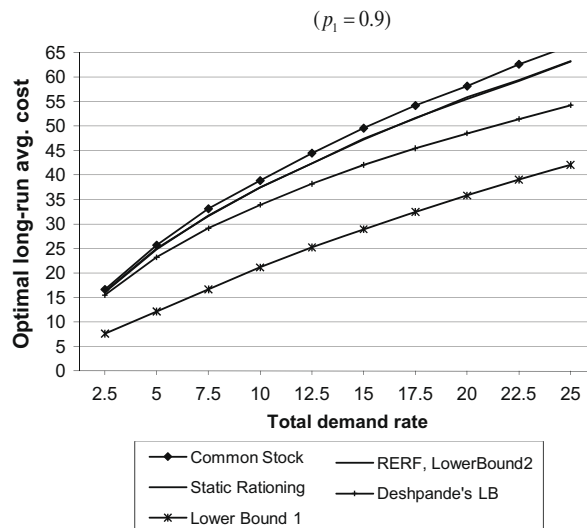**Fig. 3.** Impact of total demand rate.

Fig. 4. Impact of total demand rate.

holding and shortage costs, the relative effect of rationing diminishes as the setup cost increases. Moreover, as the batch size increases (due to the increase in the setup cost) the expected number of outstanding orders decreases and the benefit of *RERF* diminishes.

To summarize the results obtained from Table 1, we can say that rationing is a valuable tool when the total demand rate is high and the setup cost is low. In addition, if the demand rates of the customer classes are close to each other, the benefit of the static rationing over the common stock policy increases. In all such environments, *RERF* provides significant additional cost savings. Moreover, when the high portion of the demand is from class 2, the benefit of *RERF* over the static rationing policy is relatively high compared to the benefit of the static rationing over the common stock policy. Tables 2 and 3 detail the cases where *RERF* and the static policy provide high cost savings. Table 2 compares the policies for $\lambda = 25, A = 0$ and Table 3 for $\lambda = 25, A = 2$. For $A = 0$, the benefit of *RERF* over the static policy is maximized with the cost reduction of 10.10% when $p_1 = 0.1, \pi_1 = 10, \pi_2 = 2$ and $\hat{\pi}_1 = 1, \hat{\pi}_2 = 0.2$; and the benefit of the static rationing over the common stock is greatest with the reduction of 29.76% when $p_1 = 0.5, \pi_1 = 2, \pi_2 = 0.4$ and $\hat{\pi}_1 = 1, \hat{\pi}_2 = 0.2$. For $A = 2$, cost savings are maximized when $p_1 = 0.5$. In this case, 8.73% and 13.51% cost reductions are observed when $\pi_1 = 10, \pi_2 = 2, \hat{\pi}_1 = 1, \hat{\pi}_2 = 0.2$ and $\pi_1 = 2, \pi_2 = 0.4, \hat{\pi}_1 = 1, \hat{\pi}_2 = 0.2$ correspondingly for *RERF* and the static rationing policy.

As expected, cost reductions are greatest when $r_1 = \frac{\pi_1}{\pi_2}$ and $\hat{r}_1 = \frac{\hat{\pi}_1}{\hat{\pi}_2}$ are at their maximum value because service differentiation is meaningful when class 1 shortage costs are high compared to the class 2 shortage costs 2. But, in both cases *RERF* provides the maximum cost reduction over the static policy when $\pi_2 = 2$ although the static policy provides when $\pi_2 = 0.4$. This can be explained by the tendency of *RERF* to decrease the class 2 shortage costs. When the backorder cost of class 2 is high, *RERF* performs better. This fact can also be seen by comparing the cases $\pi_1 = 10, \hat{\pi}_1 = 5$ and $(\pi_1 = 2, \hat{\pi}_1 = 5)$ in Tables 2 and 3.

As discussed in the previous section, *RERF* assumes the exponential flow function $f(a_i(t)) = e^{-\lambda_1 (L-a_i(t))n}$ where n is another policy parameter to be optimized with (Q, R, K) values. In our simulation study we consider only integer values of n to obtain the optimal parameters in a reasonable amount of time. Interestingly, for both $\lambda = 25$ and $\lambda = 5, n^* \in [3, 11]$ when $p_1 = 0.1, n^* \in [1, 4]$ when $p_1 = 0.5, n^* \in [1, 3]$ when $p_1 = 0.9$, where n^* stands for the optimum value of the power n. Although n varies in small ranges, we observe that it is not effective to use an appropriate constant n value and optimize the policy only over (Q, R, K) values because changes in n results in considerable cost savings.

As a final test of performance, we compare the results of Teunter and Haneveld (2008) with the results obtained through *RERF*. Under the at-most-one-outstanding-order assumption, Teunter and Haneveld (2008) find the optimal critical remaining lead-time values. In their numerical analysis, they consider two cases. In Example 1, they assume $L = \frac{13}{24}, \lambda_1 = 0.222, \lambda_2 = 1.444, h = 1.4, \tilde{\pi}_1 = 150, \tilde{\pi}_2 = 6.5, A = 0.42$. For this case, they propose that only if the remaining lead-time is greater than 0.248, one item should be reserved for class 1 demand. The average total cost for their suggested policy is 3951. However, for the same setting the minimum average total cost obtained through *RERF* is 3805 (the optimal parameters are $(Q^*, r^*, K^*, n^*) = (2, 1, 1, 15)$). As the second example they consider a setting in which $L = 1, \lambda_1 = 4, \lambda_2 = 10, h = 1, \tilde{\pi}_1 = 100, \tilde{\pi}_2 = 10, A = 0.025$. In this setting, the authors conclude that it is optimal to reserve at most five items for class 1 demand arrivals. The reserved number decreases with the remaining lead-time. The cost of applying this policy with the optimum reorder point, order quantity and rationing times is 8777. Similar to Example 1, *RERF* provides a lower cost. With the optimal parameters $(Q^*, r^*, K^*, n^*) = (2, 19, 2, 5)$, the cost of *RERF* is 8479.

5.2. Lost sales case

We performed a similar simulation study also for the lost sales case with the same data set. Since there is no time-dependent lost sale cost, for this case we generated 72 cases varying in $\lambda, p_1 = \frac{\lambda_1}{\lambda}, \pi_1, r_1 = \frac{\pi_1}{\pi_2}$. The findings are summarized in Table 4, which is the counterpart of Table 1 for the lost sales case.

It seems that the dynamic policy does not provide savings comparable to the ones in backordering environment. The benefit of *RERF* over the static policy is greatest with the reduction of 4.30% when $\lambda = 25, A = 0, p_1 = 0.5$. As in the backordering case, the cost reduction obtained by *RERF* increases with λ and decreases with A. Similarly, the highest reductions are observed when p_1 is low or close to 0.5. On the other hand, it is not possible to say similar things to the backordering case for the behavior of the static policy. Table 4 illustrates that the

Table 4
Lost sales: comparison of policies for all possible shortage cost pairs.

		G_{RERF}		G_{SP}	
		Max	Avg.	Max	Avg.
$\lambda = 25$	$A = 0$	$p_1 = 0.10$	3.73	1.67	0.15	0.04
		0.50	4.30	2.05	8.93	3.71
		0.90	0.80	0.43	3.20	1.34
	$A = 2$	$p_1 = 0.10$	3.79	1.15	0.46	0.22
		0.50	2.68	0.86	7.04	3.22
		0.90	0.46	0.13	2.32	1.11
	$A = 10$	$p_1 = 0.10$	1.45	0.53	8.58	2.65
		0.50	2.13	0.71	6.04	2.53
		0.90	0.52	0.26	1.65	0.92
$\lambda = 5$	$A = 0$	$p_1 = 0.10$	0.47	0.31	0.08	0.02
		0.50	1.66	0.51	4.03	1.04
		0.90	0.38	0.16	1.48	0.45
	$A = 2$	$p_1 = 0.10$	0.60	0.34	0.10	0.03
		0.50	0.64	0.26	4.85	1.23
		0.90	0.43	0.15	0.83	0.34
	$A = 10$	$p_1 = 0.10$	0.25	0.16	21.27	9.96
		0.50	0.17	0.10	6.27	3.58
		0.90	0.38	0.19	1.37	0.62

static policy performs better when the setup cost increases. Here, the rationing policy is mostly effective as a demand admission control mechanism. Moreover, the highest cost reduction observed when $\lambda = 25$ and $p_1 = 0.1$. Contrary to the backordering case, inventory position does not change with the demands that are not satisfied (lost). In addition, there is no clearing issue in the lost sales case and so all the units of an arriving replenishment order are used to increase the stock. Due to these facts it is not easy to characterize the behavior of the performance of rationing strategy in the lost sales case.

6. Conclusion and extensions

In this paper, we propose a new class of dynamic rationing policies for continuous-review inventory systems with multiple customer classes. The new class of policies is based on the idea of rationing the inventory as if the outstanding replenishments were flowing into the system in a continuous fashion. The age information for all the outstanding orders is used to modify the inventory level dynamically. Upon a discussion delineating the effect of the flow function on the inventory dynamics, we suggest a policy that assumes exponential flow of the replenishment orders.

We conduct a simulation study to quantify the relative gain of our dynamic policy over the static rationing policy. We characterize the settings where the dynamic policy performs well and the rationing is a valuable tool to differentiate customers. We show that rationing is an attractive tool when total demand rate is high and setup cost is low. Furthermore, if the demand rates of the customer classes are close to each other, the benefit of the static rationing over the common stock policy increases. We demonstrate that in all these settings and moreover in the settings where the high portion of the demand is from class 2, *RERF* provides significant additional cost savings. We also propose two lower bounds on the cost of the unknown optimal dynamic policy and show that, when the demand is mostly from class 1, it is not possible to obtain any extra benefit over the static policy even with the unknown optimal policy.

Although the analytical analysis of the current form of the policy is intractable without some simplifying assumptions, considering the exponential lead-time case would be a worthwhile extension. If the lead-times are exponentially distributed, the age information of the outstanding orders would be irrelevant for the dynamic rationing decision. In that case, we should only consider the number of outstanding orders when a class 2 customer arrives. Moreover, an approximate mathematical analysis of *RERF* should be feasible under some simplifying assumptions.

Acknowledgement

References

Arslan, H., Graves, S.C., Roemer, T., 2005. A single product inventory model for multiple demand classes. Working Paper, Sawyer School of Management, Suffolk University, Boston, MA.

Atkins, D., Katircioglu, K.K., 1995. Managing inventory for multiple customers requiring different levels of service. Working Paper 94-MSC-015, University of British Columbia, Vancouver, BC, Canada.

Dekker, R., Kleijn, M.J., de Rooij, P.J., 1998. A spare parts stocking policy based on equipment criticality. International Journal of Production Economics 56, 5769–5777.

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

Deshpande, V., Cohen, M.A., Donohue, K., 2003. A threshold inventory rationing policy for service-differentiated demand classes. Management Science 49, 683–703.

Evans, R.V., 1968. Sales and restocking policies in a single inventory system. Management Science 14, 463–473.

Fadiloglu, M.M., Bulut, O., 2005. An embedded Markov chain approach to the analysis of inventory systems with backordering under rationing. Working Paper, Department of Industrial Engineering, Bilkent University, Ankara, Turkey.

Frank, K.C., Zhang, R.Q., Duenyas, I., 2003. Optimal policies for inventory systems with priority demand classes. Operations Research 5, 1993–2002.

Gayon, J.P., Vericourt, F.D., Karaesmen, F., 2006. Stock rationing in a multi class make-to-stock queue with information on the production status. Working Paper, Laboratoire Genie Industriel, 92295 Chatenay-Malabry Cedex, France.

Ha, A.Y., 1997a. Inventory rationing in a make-to-stock production system with several demand classes and lost sales. Management Science 43, 1093–1103.

Ha, A.Y., 1997b. Stock rationing policy for a make – to-stock production system with two priority classes and backordering. Naval Research Logistics 43, 458–472.

Ha, A.Y., 2000. Stock Rationing in an M/E$_k$/1 make-to-stock queue. Management Science 46, 77–87.

Kaplan, A., 1969. Stock rationing. Management Science 15, 260–267.

Kocaga, Y.L., Sen, A., 2007. Spare parts inventory management with demand lead times and rationing. IIE Transactions 39, 879–898.

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

Melchiors, P., 2003. Restricted time remembering policies for the inventory rationing problem. International Journal of Production Economics 81, 461–468.

Moon, I., Kang, S., 1998. Rationing policies for some inventory systems. Journal of Operational Research Society 49, 509–518.

Nahmias, S., Demmy, W., 1981. Operating characteristics of an inventory system with rationing. Management Science 27, 1236–1245.

Teunter, H.T., Haneveld, W.K.K., 2008. Dynamic inventory rationing strategies for inventory systems with two demand classes, Poisson demand and backordering. European Journal of Operational Research 190, 156–178.

Topkis, D.M., 1968. Optimal ordering and rationing policies in a non-stationary dynamic inventory model with n demand classes. Management Science 15, 160–176.

Veinott, A.F., 1965. Optimal policy in a dynamic, single product, non-stationary inventory model with several demand classes. Operations Research 13, 761–778.

Vericourt, F.D., Karaesmen, F., Dallery, Y., 2002. Optimal stock allocation for a capacitated supply system. Management Science 48, 1486–1501.

Zhao, H., Deshpande, V., Ryan, J.K., 2005. Inventory Sharing and rationing in decentralized dealer networks. Management Science 51, 531–547.