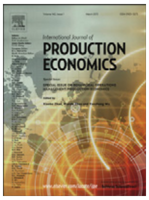
Int. J. Production Economics 144 (2013) 290–300

Contents lists available at SciVerse ScienceDirect

Int. J. Production Economics

journal homepage: www.elsevier.com/locate/ijpe



Inventory rationing decision models during replenishment lead time

Yi-Feng Hung*, Jui-Yi Hsiao

Department of Industrial Engineering and Engineering Management, National Tsing Hua University, 101 Kuang Fu Road, Sec. 2, Hsinchu 30013, Taiwan, ROC

ARTICLE INFO

Article history:
Received 4 October 2010
Accepted 8 February 2013
Available online 28 February 2013

Keywords:
Poisson process
Inventory rationing
Decision support

ABSTRACT

This study considers an inventory rationing problem in which a warehouse maintains inventory to meet various future requests that are classified into a discrete number of priorities. When there is limited inventory quantity in the warehouse, we prefer to fulfill higher priority requests over lower priority requests. When a low priority request does arrive, the warehouse may need to reject the request and reserve the inventory for later higher priority requests. However, we would rather satisfy lower priority requests than have inventory leftover when the replenishment arrives. To face uncertain request arrivals during an uncertain replenishment lead time, we develop two rationing approaches, *dynamic stochastic inventory rationing decision procedure* (DSIR) and *risk level inventory rationing decision procedure* (RLIR), to determine whether to fulfill or reject an arrival request. The simulation experiments show that RLIR provides the highest fill rate for the first priority requests among all simulated approaches, whereas DSIR provides the best overall fill rate while still maintaining a good fill rate for first priority requests. Furthermore, unlike previous studies, the proposed approaches can handle inventory systems with not only random request quantity, but also random replenishment lead time. In addition, most previous studies can only deal with two priority classes, whereas this study can handle a problem of more than two classes.

1. Introduction

Inventory includes raw materials, work-in-progress, and finished goods. The approach proposed in this study can be applied to the warehouse management of raw materials and finished goods. The purpose of inventory is to fulfill various requests by customers, especially when the acceptable waiting time set by customers is shorter than the supply lead time. Thus, warehouses should have inventory ready in stock before customer requests arrive. How to effectively manage inventory to meet demand is an important issue when the expected demand during supply lead time is more than the current on-hand inventory, especially when a major unexpected event in a supply chain disrupts the stable replenishment of inventory items.

A well-known inventory control approach, called (s, S) policy, is commonly used. When on-hand inventory is less than or equal to the re-order point s, the warehouse will issue an order to a supplier or a manufacturer to bring up the inventory position to the target level S. There is a lead time delay between the issue of an order and the receipt of the ordered items, and we call such a delay replenishment lead time or supply lead time, or in the following presentation, lead time. Also, it is common for lead time not to be a known constant parameter and to have a certain degree of uncertainty. Furthermore, the number of arrival requests and the quantity of each arrival request during the lead time are also unknown at the present time. Given that such uncertain factors exist in the inventory management problem, stock-outs might happen, especially when major unexpected events occur. Such events may cause lead time to be much longer than expected, and the demand during lead time may far exceed the original projection.

Nahmias and Demmy (1981) mentioned a practical example in the Air Force where a central supply depot maintains a shared inventory to fulfill requests from several bases that use a common spare part. However, the same item may have different request priorities. An emergency request for use in repairing an aircraft may have a higher priority than one for replenishing base stock level. When the current inventory is insufficient to fulfill the expected total quantity of all requests, the warehouse tends to first fulfill the high-priority request (i.e., the repair of an aircraft). Such a request classification is caused by the consideration of the relative importance of various demand types in the system.

For inventory items, setting a proper value for shortage cost that includes the loss of customer goodwill is difficult. The assumption of no consideration for shortage costs of different priority requests are used by a number of previous studies, such as those by Nahmias and Demmy (1981) and Haynsworth and Price (1989). The present study mainly extends their works.

---

* Corresponding author. Tel.: +886 3 5742939; fax: +886 3 5722685.
E-mail address: yifeng@ie.nthu.edu.tw (Y.-F. Hung).

0925-5273/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijpe.2013.02.017

Nahmias and Demmy (1981) provided several examples of inventory rationing problems with no consideration of shortage cost. In the present study, an additional example of the problem considered is that ambulances and shipping trucks require a common part. Determining suitable shortage costs for these two demand types may not properly reflect their relative importance. The request by an ambulance would certainly not be set aside for that by a truck. However, if shortage costs for the two priority classes are set with a ratio of 2:1, the shortage of one Priority 1 request could be exchanged to fulfill two Priority 2 requests under the same cost. In the example of the ambulance vs. truck, we would not prefer such an exchange because the availability of an ambulance is much more important than that of a truck. The cost difference between two classes in this study could be regarded as infinitely large. Supposing complete information is available on all the requests from various priority classes in the future, a lower-priority request will not be fulfilled if any higher-priority request is unsatisfied.

If requests are classified by their shortage costs in inventory rationing, the problem becomes similar to the capacity rationing problem considered by Hung and Lee (2010), who classified orders according to their profit values. Thus, the approach proposed by Hung and Lee can be applied to inventory rationing with shortage cost consideration, thereby generating similar results. The present study excludes consideration of shortage costs for inventory requests to avoid investigating a similar problem.

In our problem there is a clear priority classification among requests; that is, if we know all the requests beforehand, we should never sacrifice a high priority request to fulfill a low priority request. However, various requests will arrive randomly at different times in the future before the arrival of the replenishment. We do not have precise information on future requests; thus, we are unable to fulfill a request purely according to priority. Facing the stochastic nature of this problem, we may want to reject a request and reserve the inventory for future higher priority requests. Such an action of reservation is called “rationing.” On the other hand, if the warehouse rejects too many low priority requests and rations too much inventory for future high priority requests in the earlier stage of lead time, it might have inventory leftover when the replenishment arrives. Such leftover inventory should have been used initially to fulfill the low priority requests. How to make a correct rationing decision for inventory items is the focus of this study. The time interval in which rationing decisions have to be made is called rationing period. It starts from the time when a new order is issued and ends when the new replenishment arrives.

The performance measure used in this study is the fill rate of various priority classes. The fill rate of various priority classes is defined by the percentage of requests fulfilled by on-hand inventory. As mentioned previously, suppose that we know the information on all the requests at a single decision time; a lower priority request will never be satisfied when a higher priority one is unsatisfied. Given the information on all the requests, the optimal decision will be obtained by allocating the available inventory to the highest priority class having unfulfilled requests until there is no inventory left.

Unlike some previous studies, such as Axsater et al. (2004), Melchiors et al. (2000), and Teunter and Haneveld (2008), wherein by assigning a different unit shortage cost to each priority class, one is able to optimize a single objective of total shortage cost; the current study considers a multiple-objective solution. Higher priority requests should not be the only ones to have a higher fill rate, but lower priority requests should also have a good fill rate to avoid having inventory leftover when the new replenishment arrives. Therefore, in our experiments, both the average fill rate of the highest priority class and average overall file rate of all classes are compared among various approaches. An unsatisfied high priority request due to an inventory shortage normally occurs near the end of replenishment lead time. The shortage implies that inventory was wrongly assigned to a lower priority request at an earlier time or the highest class requests are more than originally expected.

The simplest approach to handling this problem is using the first-come-first-served (FCFS) rule, in which each arrival request is fulfilled as long as there is sufficient inventory. This approach is the easiest one and is the most commonly used in practice. However, adopting this rule sacrifices the ability to fulfill high priority requests in the future while fulfilling low priority requests now. It is inappropriate to employ this rule, especially when there is a clear indication that the demands during lead time will exceed the currently available inventory. The other extreme rule, called the highest priority only rule (HPOR), is to reject all low priority requests and fulfill only the highest priority requests. This rule will provide the highest service rate for the highest priority requests and no service at all for any lower priority requests.

Most previous studies make decisions by comparing current on-hand inventory with the *reserve level*, also called *support level* or *threshold level*. When a request arrives, if the on-hand inventory is below the reserve level, only high priority requests can be satisfied; otherwise, both low and high priority requests can be satisfied.

The works by Veinott (1965), Topkis (1968), Kaplan (1969), and Hung et al. (2012) considered inventory management problems and found reserve levels for high priority requests by minimizing defined cost functions, while Pinto (2012) maximized expected profit of the stock rationing problem in an integrated distribution system. To extend earlier studies, under the assumption of two priority classes, Nahmias and Demmy (1981) developed methods for both the periodic and continuous review of inventory systems to calculate the expected fill rate using the given values of reorder point, order quantity, and support level as specified by decision makers.

Haynsworth and Price (1989) extended Nahmias and Demmy’s work to propose a discrete time rationing policy with a desirable risk level of stock-out for high priority requests during lead time. This method utilizes a recursive backward procedure to calculate a sequence of reserve quantities for each subinterval of lead time. Ha (1997) studied a similar inventory rationing problem for a single-item, make-to-stock production system and proposed a queuing approach to find a rationing policy. Moon and Kang (1998) developed two analytical and two simulation models to improve on previous rationing methods.

The concept of rationing is also used in many different environments. Deshpande et al. (2003) considered a static rationing policy supporting two priority classes characterized by different arrival rates and shortage costs under a continuous (Q, r) review inventory framework. Ayanso et al. (2006) studied an inventory rationing problem via drop-shipping for Internet retailing. Molenaers et al. (2012) and Mohammaditabar et al. (2012) worked on the problems of inventory classification.

In this study focusing on the rationing decision within one ordering cycle, we propose two dynamic inventory rationing decision procedures—the dynamic stochastic inventory rationing decision procedure (DSIR) and the risk level inventory rationing decision procedure (RLIR). The rest of this manuscript is organized as follows. Section 2 introduces the concepts, notations, and assumptions of the proposed procedures. Section 3 discusses the two proposed procedures and then illustrates the procedures with a numerical example. Section 4 presents the results of extensive simulation experiments of various inventory rationing approaches. Finally, the conclusions are presented in Section 5.