Int. J. Production Economics 133 (2011) 393–402

Contents lists available at ScienceDirect

Int. J. Production Economics

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

An inventory reservation problem with nesting and fill rate-based performance measures

Amir-Behzad Samii [a,*], Richard Pibernik [b,c], Prashant Yadav [c,d]

[a] Operations and Technology Management Center, Vlerick Leuven Gent Management School, 9000 Gent, Belgium
[b] Supply Management Institute, European Business School, Oestrich-Winkel, Germany
[c] MIT-Zaragoza International Logistics Program, Zaragoza Logistics Center, Zaragoza, Spain
[d] MIT center for transportation and logistics (CTL), Boston, USA

ARTICLE INFO

Available online 15 April 2011

Keywords:
Inventory reservation
Service differentiation
Fill rate-based performance measures

ABSTRACT

In many inventory settings companies wish to provide customer-differentiated service levels. These may, for example, be motivated by differences in the perceived customer lifetime value or by specific contractual agreements. One approach to provide differentiated service levels is to reserve some portion of the available inventory exclusively for specific customer classes. Existing approaches to inventory reservation are typically based on the assumption that a company can assign a customer specific revenue or penalty cost to any order or unit of demand filled or unfulfilled. In practice, however, it is usually extremely difficult to accurately estimate (especially long term) monetary implications of meeting or not meeting customer demand and corresponding service level requirements. The research presented in this paper addresses the problem of setting appropriate inventory reservations for different customer classes based on fill rate-based performance measures. We model a single period inventory reservation problem with two customer classes and nesting. We develop exact expressions for two conflicting performance measures: (1) the expected fill rate of high priority customers and (2) the expected loss in the system fill rate induced by inventory reservation. With these expressions a decision maker can analyze the tradeoff between the loss in overall system performance and the higher expected fill rates for prioritized customers. We provide analytical insights into the effects of nesting and the impact of relevant problem parameters on these two performance measures. The analytical insights are illustrated and highlighted through a set of numerical examples. Although we limit our analysis to a single period inventory reservation problem, we expect that our results can be utilized in a wide range of problem settings in which a decision maker has to ration a perishable resource among different classes of customers.

1. Introduction

In many inventory settings, companies wish to provide different levels of service to different classes of customers. Customer-differentiated service levels may be motivated by differences in the perceived customer lifetime value or by specific contractual agreements that include service level guarantees. One way to provide differentiated service levels is to reserve a certain portion of the available inventory exclusively for individual classes of customers. In the relevant literature, approaches to inventory reservation are typically based on cost and revenue measures. It is assumed that a company can assign a customer specific revenue or penalty cost to any order or unit of demand filled or unfulfilled. In practice, however, it is very difficult to accurately estimate (especially the long-term) monetary implications of meeting or not meeting customer demand and corresponding service level requirements on an individual order basis. For this reason, companies commonly base inventory decisions on service level and fill rate measure targets. If a decision maker chooses to reserve inventory to ensure a higher service level for individual customer classes, he faces a decision making problem with two conflicting objectives: inventory reservation to enhance the fulfillment of demand with higher priority may have a (disproportionate) detrimental impact on the performance of the overall system. A negative impact on the overall system performance occurs if too much inventory is reserved for prioritized customers and, at the same time, the remaining inventory is not sufficient to fulfill the entire demand from customers with lower priority. As a consequence, the decision maker has to evaluate the (non-linear) tradeoff between the benefits of ensuring higher service levels for

* Corresponding author. Tel.: +32 921 09850; fax: +32 92109700.
E-mail addresses: behzad.samii@vlerick.be (A.-B. Samii), richard.pibernik@ebs.edu (R. Pibernik), pyadav@zlc.edu.es (P. Yadav).

0925-5273/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpe.2011.04.006

prioritized customer classes and the (potential) decrease in the overall performance when deciding upon reservation quantities.

The research presented in this paper attempts to lend insight into this important tradeoff. We model a single period inventory reservation problem with two classes of customers (with high and low priority) and assume that demand from both customer classes follows a Poisson process. At the beginning of the planning horizon, the decision maker decides on how much inventory to reserve exclusively for high priority customers; there will be no update of the reservation quantity during the planning horizon. We use two non-monetary performance measures to analyze the decision maker's tradeoff when determining the reservation quantity: the expected fill rate for the high priority customer class and the expected loss in the overall system fill rate induced by inventory reservation. We develop exact expressions and establish relevant properties for both performance measures. Based on the developed expressions for the individual performance measures we can fully characterize the decision maker's tradeoff dependent on the reservation level for the high priority customer class, i.e. for any reservation level we can compute the expected fill rate for the high priority customer class and the corresponding loss in the overall fill rate of the system. This information provides valuable decision support to a decision maker, not only when setting specific reservation levels, but also prior to offering certain service level agreements to a class of customers. With our expressions, the decision maker can assess whether it is prudent to "pay the price" of losing some overall system performance to achieve a certain high priority order service level.

The results of our research are distinct from prior contributions in multiple ways: first of all, we model a multi-objective inventory reservation problem based on system fill rates rather than assuming a single criterion problem based on (difficult to obtain) penalty cost and revenue measures. Secondly, we do not require strong simplifying assumptions about the arrival structure of demand from different customer classes; for example, we do not impose that demand from low priority customers arrive before demand from high priority customers (as often done in revenue management literature). Not making these simplifying assumptions adds considerable complexity to our analysis because we have to specifically account for the effects of nesting in order to model a realistic problem setting. In our setting, nesting refers to a situation where demand from the high priority customers compete equally with demand from low priority customers for any remaining unreserved inventory once (and if) the high priority customers have exhausted their reserved inventory. To individually capture the "reservation effect" and the "nesting effect" we have to explicitly account for the time structure in the demand arrival processes for both customer classes. To the best of our knowledge, this paper is the first to provide exact expressions for fill rate measures for an inventory reservation problem with nesting. Next to our analytical analysis, we conduct numerical experiments to provide additional insights into the inventory reservation problem. We illustrate the aforementioned tradeoff and show how the effects of inventory reservation (specifically the reservation and the nesting effects) depend on relevant problem parameters. We also illustrate, the conditions under which the decision maker can expect a significant negative impact on the overall system performance when reserving inventory for high priority customers. Although we limit our analysis to a single period inventory reservation problem, we expect that our results can be utilized in a wide range of problem settings in which a decision maker has to ration a perishable resource among different classes of customers.

The remainder of this paper is organized as follows: in Section 2 we review the relevant literature related to our research and highlight the contribution of our paper. In Section 3 we provide a formal characterization of the inventory reservation problem, develop exact expressions for the expected high priority fill rate and the loss in the expected overall fill rate of the system, and provide additional analytical insights. The results of a numerical analysis are presented in Section 4 and in Section 5 we summarize our findings and point to future research that may be conducted based on the results of our research.

2. Literature review

In this section we indicate the extant literature related to the research presented in this paper. A number of researchers have addressed the problem of protecting the capacity of a perishable resource for more profitable demand classes. In the literature on inventory theory, the early work of Veinott (1965) considers multiple demand classes in a multi-period, single-product setting. Topkis (1968) also solves the problem of how inventory should be allocated between demand classes. Each demand class is characterized by a different shortage cost and the allocation is based on the tradeoff between the benefit of filling demand for low class items in the current period vs. reserving the available inventory to fill higher class items in subsequent periods. Evans (1968), Kaplan (1969), and Frank et al. (2003) present models similar to the ones of Veinott and Topkis but with a different set of assumptions about the operating characteristics or customer repurchase behavior.

Nahmias and Demmy (1981) evaluate fill rates for given rationing and reorder levels in a (Q, r) inventory system with Poisson demand, two demand classes, and an inventory support level after which no low priority demand will be fulfilled. They assume that both high and low priority demand occurs at the end of the planning period as an approximation to a more realistic time structure in the random order arrival. Assuming that demand occurs at the end of the period and high priority demands are filled first will tend to underestimate high priority backorders and overestimate low priority backorders. The objective of their analysis is to develop methods for comparing fill rates when there is rationing and when there is no rationing for specified values of the reorder point, order quantity, and support level. Moon and Kang (1998), in an attempt to improve and extend the aforementioned model, develop a simulation model based on the assumption that demand occurs uniformly during the period. They claim that "…it is impossible to obtain analytical solution under this assumption". However, in this document we assume a realistic time structure of the random order arrival during the planning period without the need for a simulation-based approach and also devise the more commonly used nesting allocation policy where the reserved quantity is not a support level; i.e. the unreserved capacity is open to equal competition after the reserved quantity is consumed by the high priority orders.

On the other hand, failure to fulfill customer demand is known to affect both current cost and sales and may have a significant impact on future demand. Although measuring and mitigating the cost of this failure is not an easy task, optimal control policies and rationing levels are determined on the basis of holding and backorder cost or lost sales cost in most inventory literature. More recent examples of work in characterizing the optimal inventory rationing policy are Ha (1997), de Vericourt et al. (2002), and Duran et al. (2007) in which the objective is to minimize the expected total cost assuming a known backorder or lost sales penalty cost per unit. In parallel, Anderson et al. (2006) report the results of a sophisticated two-year long field test to measure the short- and long-run cost of a stock out in a mail-order catalog. Despite the