Int. J. Production Economics 136 (2012) 231–240

Contents lists available at SciVerse ScienceDirect

Int. J. Production Economics

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

Stock rationing under service level constraints in a vertically integrated distribution system

Roberto Pinto *

Department of Industrial Engineering, University of Bergamo, Viale Marconi 5, I—24044 Dalmine (BG), Italy

ARTICLE INFO

Article history:
Received 18 November 2010
Accepted 19 November 2011
Available online 28 November 2011

Keywords:
Rationing
Stock allocation
Service constraint
Distribution system

ABSTRACT

This paper addresses a rationing problem in a two-level, vertically integrated distribution system composed of one manufacturer and several retail points. The motivating case, developed in the vending machine sector and modeled as a newsvendor-like problem, is representative of many real settings where short-term changes in demand can be substantial while capacity modification is not a viable option. The paper provides an analytical discussion of the problem from two different standpoints: a pure profit-maximization perspective and a minimum service-level perspective, both subject to a product availability constraint that affects the service level the company can provide, and the related expected profit. By analyzing the Lagrangean formulation of the problem, we devise efficient computational procedures based on dichotomy search to find the optimal allotments to retailers, maximizing the expected profit and ensuring a minimum service level. Then, we extend the analysis to the evaluation of the highest service level that can be provided, under a product availability constraint. We identify conditions such that the proposed search procedures succeed in finding the optimal solutions, as well as bounds for the search domains. The proposed approach is legitimated under several demand distribution functions subject to a few commonly adopted restrictions that encompass many of the usually adopted continuous distributions. Finally, the paper presents a three-step decision-making framework using the proposed procedures, summarizing the decision paths the manufacturer might follow in order to optimize the allocation decision.

1. Introduction

In many business scenarios where companies face uncertain demand and building extra capacity does not represent a viable option, the available production capacity or amount of stock might not be sufficient to fulfill every possible demand realization. Though investments in capacity expansion might be recommended when long-term demand structurally exceeds capacity, from a short-term perspective the investment option could be precluded. As a result, in many supply chains the answer to shortage situations is to put the customer "on allocation," rationing capacity and stock through quantity limits, and devising proper rules to distribute the availability (Cachon and Lariviere, 1999a); uncovered demand is thus backlogged or lost, depending upon the specific sector or product.

Allocation and rationing problems[1] arise in many industrial settings, and in relation to different aspects: for example, while in a make-to-stock environment a company needs to ration the actual product inventory on hand, in a make-to-order setting it needs to ration manufacturing capacity (Hung and Lee, 2010). While the former is generally well quantified, the latter needs to be estimated in advance considering the product mix that might be requested and disruptive events that might reduce its availability once an allocation decision has already been taken.

In addition, allocation problems arise when a supplier maintains a common pool of inventory or production capacity in order to satisfy different customers, or in cases where each customer has specific contract delivery performance requirements: this exerts pressure on the supplier to differentiate stock allocation priorities (de Véricourt et al., 2002).

The objective of a company forced to manage a limited amount of a resource (capacity or stock) is to achieve the best possible result under the resource-availability constraint. A decision support tool is therefore needed in these cases, since a wrong allocation of a scarce resource may jeopardize performance. The aim is usually at balancing supply risk (the cost of having insufficient resources) and inventory risk (the cost of unsold resources).

In this paper, we address a rationing problem arising in a *vertically integrated distribution system* (or supply chain), where a company (referred to as a manufacturer or supplier) has to

* Tel.: +39 035 205 2385; fax: +39 035 205 2310.
E-mail address: roberto.pinto@unibg.it

[1] Although rationing may be viewed as a special case of the so-called allocation problem (where the former allocates all available material while the latter might allocate only a portion) (Lagodimos and Koukoumialos, 2008), we will use both terms equivalently, always assuming the need for a complete allocation.

0925-5273/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpe.2011.11.023

allocate its product availability to many retailers, which in turn sell the product to the final customers. The objective of the company is to maximize the expected profit resulting from the rationing decision.

The integration between the members of the system is realized through an organizational setting (i.e. the retail shops belong to the manufacturer), though contractual settings (i.e. return contracts) can also be considered. In both cases, the vertical integration implies that the final profit of the manufacturer is tightly linked to the performance of the retailers, which in turn depends on demand from the final customers. Thus, the rationing decision involves risk to some extent, arising from the stochastic behavior of the demand in time. A proper allocation in this context aims at avoiding situations where some retailer shops are out of stock while others are sitting on excess stock, while optimizing a profit measure.

1.1. Motivating case

The real case that motivated this research concerns a company operating in the vending machine sector; the company regularly replenishes automatic vending machines, geographically distributed on a wide area, with fresh products that have short shelf lives (salads, fruit salads, yogurt and so on). The frequency of replenishment is strongly affected by parameters such as product shelf life, magnitude of demand, location and number of installed machines, and the number of resources (agents and vans) available for replenishment. Therefore, replenishment frequency might differ quite substantially from company to company and from case to case. In the analyzed case, the product could have a very short commercial shelf life (two or three days), and the number of installed machines and available resources allows for an average rate of replenishment of once every two days. Clearly, some high-demand locations are served more frequently, while low-demand locations may receive a visit from the agents once every three or four days. For tractability purposes, we assume that all machines are visited with the same frequency.

The demand at the machines largely depends on the place where they are installed (i.e. hospitals, offices, schools, train stations, etc.) and prices can be slightly different from place to place, depending on the agreement with the contractor (the owner or the manager of the building where the machines are installed).

The company periodically faces availability problems for some products, due to suppliers' constraints and structural or contingent situations. Once the machines are replenished, the products cannot be moved from one place to another, and unmet demand is lost. Due to the short product shelf life and to the replenishment frequency, when the company's agent visits a machine either it is empty or the remaining shelf life of unsold products is too short to leave them for sale. Such unsold products are considered expired and machines can be considered empty at each agent visit.

Expired products are collected and sold at less than cost price to companies producing cattle feed or natural fertilizer, or simply disposed of. Thus, the total profit of the company depends on the selling performance of the vending machines. With regard to the vertically integrated distribution system described above, the company acts as the manufacturer, while the vending machines represent the retail shops that belongs to the manufacturer.

Such characteristics of the problem lead to the typical structure of the well-known single-period newsvendor (or newsboy) model: each vending machine represents a newsvendor whose optimal stock for each product until the next period (i.e. the next visit from the agent) may be determined considering product cost, selling price, salvage value and demand probability distribution, balancing the trade-off between lost sales and excessive stock.

The newsvendor setting presented in this paper can be easily generalized to many domains, typically seasonal products, products with long lead times, dairy or perishable products and style goods, i.e. merchandise such as fashion apparel, Christmas toys, etc., with highly uncertain demands and short selling seasons. Furthermore, the newsvendor model is distinctive of all those situations where the decision taken at the beginning of the period cannot be modified afterward, and no inventory can be carried over from one planning period to the next one.

In summary, the described problem is paradigmatic of an industry in which (1) capacity investment is sufficiently expensive that available product quantity may be less than total orders, (2) short-term changes in demand can be substantial, but short-term modifications of capacity are infeasible, (3) spot markets are not available or impose prohibitively high transaction costs and (4) the profit of the manufacturer is tightly interconnected with the performance of the retailers (Cachon and Lariviere, 1999a).

1.2. Goal and structure of the paper

Considering the contour of the general problem depicted so far, the goal of this paper is to address three specific questions about the allocation of available quantity in a vertically integrated distribution system:

1. How should a company allocate the available quantity in order to maximize the expected profit under a stochastic demand?
2. How does such allocation change if a service-level constraint is imposed at the retail level?
3. What is the highest service level that could be met at each retail shop, and what is the consequent allocation?

We provide an analytical treatment of the problem as well as answers to these questions in the form of efficient calculation methods. Furthermore, we show how these methods can be used in a decision workflow to address the rationing problem at hand.

The paper is structured as follows: Section 2 provides an overview of literature on different aspects of rationing problems. Section 3 develops the profit model that will be used in subsequent sections. Then the paper addresses the research questions; Section 4 is devoted to the maximization of the expected profit problem under availability constraints, while Section 5 extends the analysis to include inventory service-level constraints. Section 6 provides more insights into the problem of determining the highest service level that could be met at each retailer given a fixed available inventory. Section 7 sums up the different procedures proposing a use case scenario with the conjoint use of the results presented, and Section 8 draws some conclusions and delineates possible future research directions.

2. Literature review

There are many literature contributions on capacity and stock rationing. A very concise and illustrative classification can be found in Teunter and Haneveld (2008), so we only provide an abridged review of some exemplary works (including later contributions, where available) representing different aspects and classes of models and how they relate to this work.

A general and essential assumption is that it is possible to categorize the demands into classes having a different impact on the company's performance. Distinguishing between different priority classes of demand arises quite often in practice: examples can be found in airline reservations, presales of season tickets,