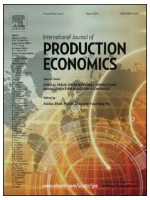
International Journal of Production Economics 185 (2017) 100–112

Contents lists available at ScienceDirect

Int. J. Production Economics

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



Optimal production and rationing policy of two-stage tandem production system

Jianjun Xu[a,d,*], Alejandro Serrano[a,b], Bing Lin[c]

[a] Zaragoza Logistics Center, Spain
[b] IESE Business School, Spain
[c] Jiangsu Normal University, China
[d] International Business College, Dongbei University of Finance and Economics, Dalian, China

ARTICLE INFO

Keywords:
Production/inventory control
Stock rationing
Tandem system
Dynamic programming
Bulk production

ABSTRACT

The rationing literature has been so far oblivious to the fact that sellable items may not only be finished products, but also intermediate products to manufacture subsequent sellable products. We address this gap by considering an inventory rationing problem in a two-product tandem make-to-stock production/inventory system. Bulk demand arrives to a partial-batch production system with exponentially-distributed production time for each batch. The management has to decide whether to run or stop production and whether the various classes of demand for both products—intermediate product (IP) and finished product or end product (EP)—have to be satisfied from available inventory or not—in which case demand is lost—in order to maximize the firm's expected profit. We present the corresponding dynamic programming formulation and characterize the optimal policy. The resulting policy depends on dynamic switching curves, which define a) thresholds to continue or discontinue production, and b) thresholds to satisfy or turn down incoming orders. We identify the key value drivers and compare various heuristic policies through extensive numerical analyses.

1. Introduction and related literature

The last decades have witnessed two economic and business trends that are hard to reconcile. First, variety of demand from customers has triggered product proliferation, which has dramatically increased the number of stock keeping units (SKU) that firms offer their customers. Second, pressure from the financial realm to reduce working capital has led manufacturing and retailing firms to—other things being equal—reduce their inventory levels. One straightforward way to achieve the latter by ignoring the former is simply by reducing the number of SKU offered. For instance, the CEO of Adidas, the second largest sportswear manufacturer in the world, announced in 2012 his decision to reduce the firm world-wide assortment by 25%, that is, 11,724 SKU out of 46,897 available SKU. The other available approach is to maintain a relatively large assortment and assume a reduction of customer service level. When that is the case, a key point to address is how to maximize the firm utility given the relative inventory scarcity. The usual approach in economics is to try to sell to those customers who are willing to pay a higher premium for the scarce available products. This can mainly be achieved through two dynamic approaches, either by modifying the price of goods (as in revenue management settings) or by turning down certain customer demand classes leading to relatively low margins when demand of high-margin classes is expected to occur in the near future (as in rationing settings). Both revenue management and rationing techniques have naturally attracted attention in academia in recent times. In the latter case, the extant operations management (OM) literature has dealt with rationing for finished products, i.e., it has assumed that available products being sold are not further transformed to obtain subsequent sellable products. However, this is not always the case. A manufactured product may well be both a finished good to be sold to outside customers and an intermediate product necessary to manufacture a finished product. Such a setting, usually referred to as a tandem production/inventory system, is common in several industries, such as pharmacy, textile, or electronics.

Consider for example the case of a large Chinese pharmaceutical manufacturer that produces and sells erythromycin thiocyanate, a key material to produce erythromycin oxime, which in turn is used to manufacture macrolide antibiotics, such as azithromycin, roxithromycin, or clarithromycin. The firm holds some inventory of erythromycin thiocyanate (IP) and erythromycin oxime (EP).[1] A portion of IP is sold

---

* Corresponding author.
E-mail address: jxu@zlc.edu.es (J. Xu).

[1] Here we use the terms intermediate and end product to refer to erythromycin thiocyanate and erythromycin oxime respectively, regardless of the fact that the latter is also a necessary material to obtain macrolide antibiotic.

http://dx.doi.org/10.1016/j.ijpe.2016.12.025
Received 13 July 2015; Received in revised form 15 December 2016; Accepted 16 December 2016

to outer factories that produce erythromycin oxime, while other portion is stored to satisfy the demand of its internal portfolio. The demand for this IP comes from various types of customers, some of which have long-term contracts in place, whereas others order only occasionally. Therefore, the allocation of inventory or rationing problem appears, especially during the selling season with limited inventory and production capacity. As described, the rationing problem for the firm has two dimensions: first, it must decide how much IP to allocate to produce EP so as to satisfy the demand from outer factories that produce macrolide antibiotic products; second, it must also decide whether outside demand for IP must be satisfied or not, taking into consideration the fact that customers possibly differ in their service level requirements thus the firm may have different shortage costs for different customers.

In the textile industry, a typical manufacturer may produce cotton threat and cloth, the former being a key component to produce the latter. Manufacturers supply both cotton threat (IP) and cloth products (EP) to different types of customers, possibly with different shortage costs and required service levels. Inventory allocation should be such that the internal demand for IP is not compromised.

The OM literature has addressed rationing settings and tandem production/inventory systems, but in isolation. As for rationing, the extant literature is extensive. In periodic-review model formulations, Veinott (1965), Topkis (1968), Cohen et al. (1988), Frank et al. (2003), Pibernik and Yadav (2009), Chen et al. (2010), Lee et al. (2011), Chew et al. (2013) study the rationing policy for a dynamic single-product inventory system with several demand classes. Wang and Tang (2014) consider a dynamic rationing policy for multiple demand classes with a mixture of lost sales and backorders. They show that rationing levels often exhibit different patterns between the priority switching points. Turgay et al. (2015) investigate a finite-horizon, discrete-time, single-product rationing problem where demand and production rates are uncertain.

Nahmias and Demmy (1981) and Deshpande et al., (2003) consider a continuous-review model of a stock rationing system. Gerchak et al. (1985) also give insightful results of the policy of inventory rationing and production in the discrete-time case.

In the make-to-stock queue settings, Ha (1997a) formulates the single-item multi-class stock rationing problem with lost sales. Optimal policy is characterized with critical numbers called rationing levels. Ha (1997b) considers stock rationing with backordering under the presence of two priority demand classes. Ha (2000) further extends previous results to the case of Erlang distributed processing time. de Vericourt et al. (2002) consider an extension of Ha (1997b) with multiple-demand class. Huang and Iravani (2008) and Xu et al. (2010) further extend the de Vericourt et al. (2002) model to more general settings, such as multiple demand classes and batch arrival. Benjaafar et al. (2010) study a production/inventory system with a single product and two customer classes where both backorders and lost sales are permitted. Enders et al. (2014) model a single-item inventory system with a high priority lost sales customer class and a lower priority backordering class. They propose a critical level (CL) policy and develop an efficient procedure to determine the average performance of a given CL policy. Isotupa (2015) analyzes a lost-sales inventory system with two classes, and shows that there is a sub-optimal policy under certain conditions. Unlike all the above papers that only consider end products, we investigate a two-stage tandem system that consists of demands from both end products (EP) and intermediate products (IP). Although in general a critical level policy has been shown to be efficient for the single-stage problems from the above papers, our results show that, in a two-stage tandem system, the optimal policy is a switch curve policy. The results from a critical level policy could be significantly deviated from the optimal one.

The topic of tandem systems has also been extensively studied by, e.g., Rosberg et al. (1982) and Weber and Stidham (1987), among others. They study make-to-stock production/inventory systems and their optimal control. Veatch and Wein (1994) find conditions under which certain simple controls are optimal using stochastic coupling arguments. Optimal results are computed using dynamic programming and compared with kanban and buffer control mechanisms, as well as with the base stock mechanism. Lee and Zipkin (1992) study a tandem-queue model where intermediate and finished goods can be produced and stored in advance of demand. They propose a tractable approximation scheme. Duenyas and Patana-Anake (1998) consider an N-stage tandem manufacturing system that produces a single product. They analyze the performance of base-stock policies. Simulation results indicate that the proposed base-stock policies perform very well. Wang et al. (2013) model a two-stage hybrid manufacturing/service system, in which one factory manufactures products in the first stage and then one service center realizes product-service systems by adding some product-based services on the products in the second stage, with both demands for product and product-service systems. None of these papers study rationing problems.

One closely related paper to ours is Ceryan et al. (2012) who consider production/inventory control for a two-stage system where intermediate components are produced and an end product is assembled from these components. The structure of optimal policies for demand admission, component production and product assembly decisions are characterized. They also extend the model to the case with multiple customer classes. In contrast to that paper, which considers a mixed make-to-stock and assembly-to-order system, our model focuses on a two-stage make-to-stock system. Further details are given in Section 3 after we present our analytical results.

While papers in the make-to-stock queue stream assume unit production, we relax this constraint by assuming that the facility can produce all units in a batch up to its capacity concurrently. Batch production is common in the pharmaceutical industry (Leuenberger, 2001), and is also common in bakeries and in the manufacture of sports shoes, purifying water, inks, paints and adhesives. Following Gross and Harris (1998), we allow for bulk demand and production in partial batches. Specifically, each facility can process partial batches up to a maximum size. If fewer units than the maximum capacity are in the production process, new arrivals immediately enter production up to the maximum size limit, and finish together with the already in-process units, regardless of the entry time into production. The amount of time required for the service of any batch is an exponentially distributed random variable, whether or not the batch is of the full size. Such setting is referred by Gross and Harris (1998) as a bulk-service system $M/M^{[Y]}/1$. Besides applications in production systems, bulk-service systems also have many applications in transportation—Kahraman and Gosavi (2011), Nemhauser and Yu (1972), Powell and Humblet (1986), service operations, computer and telecommunication—Germs and van Foreest (2013).

In addition to the bulk-production and batch-demand extension, we also contribute to the make-to-stock literature in the following ways. First, we generalize the usual setting in rationing systems by considering precisely a tandem production/inventory system. In other words, we play the rationing game for different classes of demand not only for a sellable IP, but also for the different classes of demand for a subsequent EP, made from the corresponding IP. We show the optimality of monotone switching curves for production and rationing. Second, in addition to stock rationing, we introduce production capacity rationing when multiple production modes are available. We claim that our main results can be extended to this case and production capacity rationing policy can be characterized by additional monotone switching curves. Third, we conduct a sensitivity analysis to gain insight on how the objective function and the switching curves change with the relevant parameters. Furthermore, we investigate the value of rationing using numerical experiments. Our results show that with a desirable customer segmentation and rationing policies, the firm could increase its revenue by more than 3% on average. We further compare two types of rationing in the system and test the effectiveness of critical