The joint stock and capacity rationings of a make-to-stock system with flexible demand

Lin, Bing; Chen, Shaoxiang; Feng, Yi; Xu, Jianjun

2018

Lin, B., Chen, S., Feng, Y., & Xu, J. (2018). The joint stock and capacity rationings of a make-to-stock system with flexible demand. *Asia-Pacific Journal of Operational Research*, 35(01), 1850004-1-27. doi:10.1142/S0217595918500045

<https://hdl.handle.net/10356/89851>

<https://doi.org/10.1142/S0217595918500045>

© 2018 World Scientific Publishing & Operational Research Society of Singapore. All rights reserved. This paper was published in *Asia-Pacific Journal of Operational Research* and is made available with permission of World Scientific Publishing & Operational Research Society of Singapore.

Downloaded on 26 Apr 2023 22:59:38 SGT

The Joint Stock and Capacity Rationings of a Make-To-Stock System with Flexible Demand

Bing Lin*

*Department of Logistics, School of Business
Jiangsu Normal University, Xuzhou 221116, P. R. China
linbing@pmail.ntu.edu.sg*

Shaoxiang Chen

*Division of Information Technology and Operations Management
Nanyang Business School, Nanyang Technological University
Nanyang Avenue 639798, Singapore
aschen@ntu.edu.sg*

Yi Feng

*School of Management and Economics
University of Electronic Science and Technology of China
Chengdu 610054, P. R. China
fengyi@uestc.edu.cn*

Jianjun Xu

*International Business College and Institute of Supply Chain Analytics
Dongbei University of Finance & Economics
Dalian 116025, P. R. China
xujianjun@dufe.edu.cn*

Received 14 May 2016

Revised 15 January 2017

Accepted 17 October 2017

Published 15 January 2018

This paper studies the joint stock and production capacity rationing policies for a make-to-stock system with two partially substitutable products manufactured at two factories, respectively. Customers are classified into three types with each type further segmented into multiple classes. Type I and Type II customers are selective and accept their favorite products only. Type III accepts either of the two products and hence its demand is regarded as the flexible demand. The management has to decide whether to run or stop production and whether the various classes of demand for each type of customers have to be satisfied from available inventory or not — in which case demand is lost — in order to minimize the firm's inventory and production costs. We formulate the problem as a make-to-stock queue system and characterize the optimal joint stock and production rationing

*Corresponding author.

policies as monotone switching curves. Numerical examples are given to illustrate the optimal policies and demonstrate the value of flexible demand. Moreover, we develop a simple heuristic policy and compare the costs associated with the optimal and heuristic policies for different parameters.

Keywords: Production; inventory; rationing; dynamic programming; Markov decision processes.

1. Introduction

The existence of heterogeneous demand is a common phenomenon in a wide range of industries. Some customers (e.g., who often place large orders) may be considered as more important than others (Teunter and Klein Haneveld, 2008). In addition, many suppliers wish to provide different levels of service to different customers. The higher priority a customer has, the higher penalty cost will incur if the customer is rejected due to the stock shortage. Thus, a common practice for balancing supply with heterogeneous demand is to categorize customers into different priority classes based on their economic values (Arslan *et al.*, 2007), and ration inventories among different demand classes. That is, for customers who are economically valuable to the suppliers, it is necessary to reserve stock for them when the production capacity is limited.

In addition to economic values to the suppliers, in this paper we also consider the other variable that is used to segment customers: customer flexibility and customer selectivity. Iravani *et al.* (2003) define customer flexibility and customer selectivity as the customers' willingness to substitute or ignore some important features of his original order. Based on customers' willingness to substitute, customers can be categorized as selective customers, who are not going to buy a product unless that product has a set of specific features (key items), and flexible customers, who are willing to compromise his key items. Since it is common in practice that suppliers face both selective and flexible customers, a rationing question then naturally arises: in addition to the rationing among different demand classes, how to allocate inventories between selective and flexible customers?

Besides stock rationing that adopts the marketing strategy of customer segmentation, another common approach to balance supply with demand, called capacity rationing, is to make a firm's production more flexible by introducing multiple production modes, such as subcontracting. Management can flexibly adjust the production capacity, e.g., when demand is low, a normal production mode is taken; when demand is high, overtime production and subcontracting are taken. With flexible production options, a manager can respond quickly to incoming demand by adjusting the production capacity. The key idea of capacity rationing is to decide whether to run or stop production and provide volume flexibility which is the ability of a manufacturing system to operate profitably at different overall output levels (Sethi and Sethi, 1990). According to Goyal and Netessine (2011), volume flexibility allows a firm to profitably increase or decrease production above and below the installed capacity.

In this paper, we analyze both stock and production capacity rationing controls for systems that have both selective and flexible customers from multiple demand classes that are classified by their economic values. More specifically, we consider a make-to-stock system with factories A and B to manufacture two products, respectively. Factory A (B) has $n_1(n_2)$ production modes, yielding $n_1(n_2)$ classes of production rates. Customer demands are classified into three types, say, Type I, II and III. Type I customers are only satisfied by factory A. Type II customers are only satisfied by factory B. Type III is satisfied by either factory A or B. Each customer type has m priority classes. Upon arrival, demands are either satisfied or denied and unmet demands are lost. The problem is illustrated in Fig. 1.

Our model has many potential applications. For instance, a floorboard-making company has two factories with one factory making floorboard from bamboos, and the other making it from woods. Some customers choose bamboo-made floorboard specifically, some choose wood-made floorboard specifically, and the rest do not care about which material the floorboard is made from. Then it is up to the manufacturer to decide which item is assigned to the flexible customers. Another application is in the manufacturing network. The firm has two manufacturing bases in China, one in Yangtze River Delta (YRD) area, the other in Pearl River Delta (PRD) area. Each manufacturing base can employ different production rates associated with different plants. The final products manufactured in YRD will be mainly sent to the North American market, and final products in PRD will be mainly sent to the European market. Both products could be sent to the countries in the Asia-Pacific area. It needs better coordination between the two manufacturing bases to achieve an efficient usage of production capacities and inventories.

Stock rationing problems have been extensively studied in literature. For the single-item inventory systems under periodic review, Topkis (1968) and Veinott (1965) are among the earliest ones to study the inventory management problems with multiple demand classes. Cohen *et al.* (1988) derive a (s, S) periodic

Fig. 1. The make-to-stock system with three types of demand.

review policy for the lost-sales problem with two demand classes. In Frank *et al.* (2003), the rationing policy has been examined for single-item inventory system with two demand classes, one deterministic and the other stochastic. Furthermore, Chen *et al.* (2010) give a partial characterization of the optimal joint ordering and rationing policy for a backorder inventory system with two demand classes. More recently, Lee *et al.* (2011), Chew *et al.* (2013), and Liu *et al.* (2015) consider the stock rationing of the single-item inventory system with multiple demand classes.

For the single-item inventory system under continuous review, Arslan *et al.* (2007), de Vericourt *et al.* (2002), Deshpande *et al.* (2003), Ha (1997a,b, 2000), Nahmias and Demmy (1981) consider the rationing and production policies under different scenarios and assumptions. Benjaafar and ElHafsi (2006) consider production and inventory controls of an assemble-to-order system with multiple demand classes. Axsäter *et al.* (2004) and Huang and Iravani (2006) study the two-echelon inventory rationing problem. Zhao *et al.* (2005) investigate an inventory rationing problem in a network setting. Fadiloglu and Bulut (2010) propose a dynamic rationing policy for the continuous-review inventory systems based on the status of outstanding orders. Moreover, both Huang and Iravani (2008) and Xu *et al.* (2010) consider stock rationing problems with batch demand while Pang *et al.* (2014) study inventory rationing in the make-to-stock systems with batch ordering and lost sales.

Recently, some rationing problems are of special interest. For instance, Papier and Thonemann (2010) study capacity rationing for stochastic rental systems with advance demand information. Bulut and Fadiloglu (2011) characterize the optimal production and stock rationing policies of a make-to-stock system with parallel production lines. Wang and Tang (2014) consider a dynamic rationing problem for the case with a mixture of lost sales and backorders. Enders *et al.* (2014) study a problem in which customers are distinguished by whether or not they are willing to wait for backordered demand, i.e., one customer class leaves when demand is not satisfied immediately, while the other customer class is willing to wait and demand is backordered. Turgay *et al.* (2015) study a discrete-time finite-horizon inventory rationing problem with uncertain demand and production rates. Xu *et al.* (2017) characterize the optimal rationing policy for a tandem make-to-stock system and provided motivated examples for the real-life applications of multiple demand classes in manufacturing industry.

Our research also relates to capacity rationing and volume flexibility. Rajagopalan and Soteriou (1994) consider a model addressing capacity acquisition, disposal and replacement problems with discrete equipment sizes. They point out that capacity may often be disposed due to declining demand or replaced due to deterioration or obsolescence. They present linear programming formulations and identify special structures. Tomlin (2006) considers a single-product model in which a firm can either source from an unreliable supplier, or a reliable supplier who has volume flexibility (which is defined as the amount and speed with which extra capacity becomes available). Tomlin shows that volume flexibility can substantially benefit the firm, as it enables contingent rerouting to be an element of the firm's

strategy, and this can significantly reduce the firm's costs. Donohue (2000) analyzes a set of supply contract used when a manufacturer operates two modes: the first production mode is cheap but requires a long lead time while the second is expensive but offers quick turnaround. Different with these papers, we study the problem in a make-to-stock queue setting and our focus is on production rationing decisions that dynamically identify the optimal production mode.

The research on how to allocate capacity between selective and flexible customers is rare. Irvani *et al.* (2003) consider an assemble-to-order system in which each customer order consists of a mix of key and non-key items. They assume that some proportion of customers accept substitutions for missing items. Their analysis incorporates customer preferences and flexibility. They obtain the performance measures of the assemble-to-order system and introduce new measures which evaluate different levels of customer satisfaction. In this paper, however, we focus on stock rationing policy between flexible customers and selective customers. To the best of our knowledge, this paper is the first to integrate stock rationing among customers in terms of both their selectiveness and economic values, and capacity rationing in the sense that production capacity can be adjusted based on demand information.

The rest of the paper is organized as follows. In Sec. 2, we present the dynamic programming model of the problem. In Sec. 3, the optimal policy has been characterized and an extension to the batch demand case is also discussed. The value of flexible demand is studied in Sec. 4. Numerical analysis illustrates our results in Sec. 5. We draw the conclusions in Sec. 6.

2. Model Formulation

For the convenience of readers, we first list some major notations for model formulation as follows:

- i : index for the classes of demand
- l : index for the types of demand
- j : index for the classes of production rate at factory A
- k : index for the classes of production rate at factory B
- λ_{li} : the arrival rate for demand of type l and class i
- $\mu_{1j}(\mu_{2k})$: the product rate of class j at factory A (of class k at factory B)
- π_{li} : the unit penalty cost for lost sales of type l and class i
- $c_{1j}(c_{2k})$: the unit product cost of class j at factory A (of class k at factory B)
- $h(x_1, x_2)$: the inventory holding cost function
- $V(x_1, x_2)$: the optimal cost function.

Let the state indicated by $x = (x_1, x_2)$, where x_1 and x_2 are the respective inventory levels of two factories, which take nonnegative integer values. In our model

formulation, it is reasonable to assume that Type III demand is lost if it cannot be satisfied by either product A or B. Analogous to the study of many other related papers on inventory rationing, Type I and Type II demand are also assumed to be in the lost-sales case. In fact, lost-sales case is often the only option for the reasons of market competition, special product category, long lead time and operating policy.

According to the above assumptions, x_1 and x_2 are restricted to be nonnegative integers, i.e., $x = (x_1, x_2) \in Z_+^2$, where the state space Z_+^2 is a two-dimension non-negative integer set. We assume that $h(x_1, x_2) = h_1x_1 + h_2x_2$, where h_1 and h_2 are the unit holding costs per unit time at factories A and B, respectively.

Assume that production time follows an exponential distribution. Suppose each factory has multiple classes of production rates, denoted by μ_{1j} and μ_{2k} ($j = 1, \dots, n_1$, $k = 1, \dots, n_2$), respectively, and each production rate is associated with a production mode. Unit production cost associated with the production rates $\mu_{1j}(\mu_{2k})$ are $c_{1j}(c_{2k})$. Without loss of generality, we assume that $c_{11} < c_{12} < \dots < c_{1n_1}$ ($c_{21} < c_{22} < \dots < c_{2n_2}$) for factory A (factory B). Each demand type has m priority classes, which differ in their penalty costs for lost sales denoted by π_{1i}, π_{2i} and π_{3i} ($i = 1, \dots, m$), respectively. Without loss of generality, let $\pi_{l1} > \pi_{l2} > \dots > \pi_{lm}$, ($l = 1, 2, 3$), i.e., the priorities of customers are in the descending order with class 1 being the highest one. Furthermore, we assume the production costs are no more than the penalty costs, which include the potential revenue losses. Suppose each class of demand arrives in accordance with a Poisson process and the arrival rates are $\lambda_{1i}, \lambda_{2i}$ and λ_{3i} ($i = 1, \dots, m$), respectively. We further assume that $\sum_{i=1}^m (\lambda_{1i} + \lambda_{2i} + \lambda_{3i}) < \sum_{j=1}^{n_1} \mu_{1j} + \sum_{k=1}^{n_2} \mu_{2k}$, which implies that total production capacity can handle total demand in the long run.

Let $e_1 = (1, 0)^T$, $e_2 = (0, 1)^T$, and $0 < \alpha < 1$ be the continuous-time discount rate. Given the initial state $x = (x_1, x_2)$, the expected total discounted cost $J_u(x)$ obtained for a policy u is:

$$J_u(x) = E_x^u \left[\int_0^\infty e^{-\alpha t} h(x) dt + \sum_{l=1}^3 \sum_{i=1}^m \int_0^\infty e^{-\alpha t} \pi_{li} dM_{li}(t) + \sum_{j=1}^{n_1} \int_0^\infty e^{-\alpha t} c_{1j} dN_{1j}(t) + \sum_{k=1}^{n_2} \int_0^\infty e^{-\alpha t} c_{2k} dN_{2k}(t) \right],$$

where $M_{li}(t)$, $N_{1j}(t)$ and $N_{2k}(t)$ represent, respectively, up to time t , the number of rejected customers from the i th class and the l th type, the number of items produced with the j th production rate at factory A, and the k th production rate at factory B. Hence, on the right-hand side of the above objective function, the first term in the square bracket is the discounted holding cost. The second is the discounted penalty cost incurred by rejecting demand. The third and the fourth terms are the discounted production costs of two factories, respectively. Then the optimal cost function $V(x)$ is the minimum of $J_u(x)$ over the stationary policy $u \in U$, where U

is the set of all admissible policies. Mathematically,

$$V(x) = \min_{u \in U} J_u(x).$$

Following the uniformization technique in Lippman (1975), let

$$\Lambda = \sum_{i=1}^m (\lambda_{1i} + \lambda_{2i} + \lambda_{3i}) + \sum_{j=1}^{n_1} \mu_{1j} + \sum_{k=1}^{n_2} \mu_{2k}.$$

After the uniformization of state transitions, the original continuous-time stochastic system is observed at evenly spaced points in time, which corresponds to either a demand arrival event or the event of the production completion of one item. When a stationary policy u is employed, stocking rationing decisions $u_{1i}^d \in u$ can take values 0 and 1, indicating the decision of rejecting or satisfying the i th class of Type I demand by factory A, respectively; $u_{2i}^d \in u$ can take values 0 and 2, indicating the decision of rejecting or satisfying the i th class of Type II demand by factory B, respectively; $u_{3i}^d \in u$ can take values 0, 1, and 2, indicating the decisions of rejecting, or satisfying the i th class of Type III demand by factory A, or by factory B, respectively. The production capacity rationing decisions $u_{1j}^p \in u$ can take values 0 and 1, indicating production idleness and usage of the j th class production rate at factory A, respectively; $u_{2k}^p \in u$ can take values 0 and 2, indicating production idleness and usage of the k th class production rate at factory B, respectively.

Let the state after transition is y . Given a state (x_1, x_2) and for each control value of u_{1i}^d , u_{1j}^p and u_{2k}^p , define the transition probability function $p(y | x, u)$ as

$$\begin{aligned} p(y | x, u) = & \sum_{i=1}^m \frac{\lambda_{1i}}{\Lambda} I\{x - e_1 | x, u_{1i}^d = 1\} + \sum_{i=1}^m \frac{\lambda_{3i}}{\Lambda} I\{x - e_1 | x, u_{3i}^d = 1\} \\ & + \sum_{i=1}^m \frac{\lambda_{2i}}{\Lambda} I\{x - e_2 | (x_1, x_2), u_{2i}^d = 2\} + \sum_{i=1}^m \frac{\lambda_{3i}}{\Lambda} I\{x - e_2 | x, u_{3i}^d = 2\} \\ & + \sum_{i=1}^m \frac{\lambda_{1i}}{\Lambda} I\{x | x, u_{1i}^d = 0\} + \sum_{i=1}^m \frac{\lambda_{2i}}{\Lambda} I\{x | x, u_{2i}^d = 0\} \\ & + \sum_{i=1}^m \frac{\lambda_{3i}}{\Lambda} I\{x | x, u_{3i}^d = 0\} + \sum_{j=1}^{n_1} \frac{\mu_{1j}}{\Lambda} I\{x + e_1 | x, u_{1j}^p = 1\} \\ & + \sum_{k=1}^{n_2} \frac{\mu_{2k}}{\Lambda} I\{x + e_2 | x, u_{2k}^p = 2\} + \sum_{j=1}^{n_1} \frac{\mu_{1j}}{\Lambda} I\{x | x, u_{1j}^p = 0\} \\ & + \sum_{k=1}^{n_2} \frac{\mu_{2k}}{\Lambda} I\{x | x, u_{2k}^p = 0\}, \end{aligned}$$

where $I\{\cdot\}$ is the indicator function for a transition. To ease notations in each indicator function, we do not write down other control actions which all equal

Fig. 2. The state transition diagram with a control policy that all Type I demands are accepted, Types II and III are rejected, and all production capacities at each factory are exploited.

to 0 except for the one that makes the transition to happen. Then, in Fig. 2, we show the state transitions for a special case. For simplicity, we only draw the state transition diagram with a control policy that all classes of Type I demand are accepted, Types II and III are rejected, and all production capacities at each factory are exploited. For shorthand notations, let $\lambda_1 = \sum_{i=1}^m \lambda_{1i}$, $\lambda_2 = \sum_{i=1}^m \lambda_{2i}$, $\lambda_3 = \sum_{i=1}^m \lambda_{3i}$, $\mu_1 = \sum_{j=1}^{n_1} \mu_{1j}$ and $\mu_2 = \sum_{k=1}^{n_2} \mu_{2k}$. Thus, $\Lambda = \lambda_1 + \lambda_2 + \lambda_3 + \mu_1 + \mu_2$.

Next, define the operators $T_{1,i}$, $T_{2,i}$, $T_{3,i}$, $T_{4,i}$, $T_{5,i}$ as follows:

$$T_{1,i}V(x) = \begin{cases} \min\{V(x - e_1), V(x) + \pi_{1i}\} & \text{if } x_1 > 0 \\ V(x) + \pi_{1i} & x_1 = 0 \end{cases},$$

$$T_{2,i}V(x) = \begin{cases} \min\{V(x - e_2), V(x) + \pi_{2i}\} & \text{if } x_2 > 0 \\ V(x) + \pi_{2i} & x_2 = 0 \end{cases},$$

$$T_{3,i}V(x) = \begin{cases} \min\{V(x - e_1), V(x - e_2), V(x) + \pi_{3i}\} & \text{if } x_1 > 0 \text{ and } x_2 > 0 \\ \min\{V(x - e_1), V(x) + \pi_{3i}\} & \text{if } x_1 > 0 \text{ and } x_2 = 0 \\ \min\{V(x - e_2), V(x) + \pi_{3i}\} & \text{if } x_2 > 0 \text{ and } x_1 = 0 \\ V(x) + \pi_{3i} & x_1 = 0 \text{ and } x_2 = 0 \end{cases},$$

$$T_{4,j}V(x) = \min\{c_{1j} + V(x + e_1), V(x)\},$$

$$T_{5,k}V(x) = \min\{c_{2k} + V(x + e_2), V(x)\},$$

where $T_{1,i}$, $T_{2,i}$, $T_{3,i}$ are the operators for admission controls of the i th class demand of Type I, Type II and Type III, respectively and $T_{4,j}$, $T_{5,k}$ are for the production

control of factories A and B, respectively. Define T as

$$TV = \frac{\Lambda}{\alpha + \Lambda} \left\{ \frac{h}{\Lambda} + \sum_{i=1}^m \frac{\lambda_{1i}}{\Lambda} T_{1,i} V + \sum_{i=1}^m \frac{\lambda_{2i}}{\Lambda} T_{2,i} V \right. \\ \left. + \sum_{i=1}^m \frac{\lambda_{3i}}{\Lambda} T_{3,i} V + \sum_{j=1}^{n_1} \frac{\mu_{1j}}{\Lambda} T_{4,j} V + \sum_{k=1}^{n_2} \frac{\mu_{2k}}{\Lambda} T_{5,k} V \right\}.$$

Then, after rescaling time to achieve $\alpha + \Lambda = 1$,

$$TV = h + \sum_{i=1}^m \lambda_{1i} T_{1,i} V + \sum_{i=1}^m \lambda_{2i} T_{2,i} V + \sum_{i=1}^m \lambda_{3i} T_{3,i} V \\ + \sum_{j=1}^{n_1} \mu_{1j} T_{4,j} V + \sum_{k=1}^{n_2} \mu_{2k} T_{5,k} V. \quad (1)$$

According to Puterman (1994), for the discount factor $\frac{\Lambda}{\alpha + \Lambda} < 1$, T is a contraction mapping and V satisfies the Bellman equation:

$$V = TV. \quad (2)$$

3. Characterization of the Optimal Policy

3.1. The switching-curve policy

In this subsection, we explore the structural properties of the optimal policy. First, define two differences: $\Delta_1 V(x) = V(x + e_1) - V(x)$, $\Delta_2 V(x) = V(x + e_2) - V(x)$. Then, define $\hat{V}$ as the set of functions on Z_+^2 such that if $V \in \hat{V}$,

- (a) $\Delta_1 V(x) \uparrow x_1 \uparrow x_2$;
- (b) $\Delta_2 V(x) \uparrow x_1 \uparrow x_2$;
- (c) $\Delta_1 V(x) - \Delta_2 V(x) \uparrow x_1 \downarrow x_2$;

where $\uparrow$ and $\downarrow$ denote non-decreasing and non-increasing, respectively. $\Delta_1 V(x) \uparrow x_1$ in (a) and $\Delta_2 V(x) \uparrow x_2$ in (b) refer to the discrete convexity in x_1 and x_2 , respectively, $\Delta_1 V(x) \uparrow x_2$ in (a) and $\Delta_2 V(x) \uparrow x_1$ in (b) are the supermodular property of $V(x)$ and (c) refers to the difference monotone as defined in Chen (2004). For brevity, we use increasing and decreasing to represent non-decreasing and non-increasing respectively. To characterize the optimal policy, we need the following lemma.

Lemma 1. $T_{1,i} V, T_{2,i} V, T_{3,i} V, T_{4,j} V, T_{5,k} V$ and $TV \in \hat{V}$ if $V \in \hat{V}$.

Proof. The proof is lengthy and readers are referred to Lin (2008) for details. $\square$

Lemma 1 states that the operators preserve the structure properties (a), (b) and (c). In the sequel we can prove the following lemma.

Lemma 2. *The optimal cost function $V \in \hat{V}$.*

Proof. Please see the appendix. □

Next, we prove the existence of an optimal policy for the model.

Lemma 3. *There exists an optimal policy for the Bellman equation (2).*

Proof. Please see the appendix. □

To further investigate the optimal policy, we define the switching functions:

$$S_1(x_2, i) = \min\{x_1 \mid \pi_{1i} + V(x) - V(x - e_1) \geq 0, \text{ given } x_2\},$$

$$S_2(x_1, i) = \min\{x_2 \mid \pi_{2i} + V(x) - V(x - e_2) \geq 0, \text{ given } x_1\},$$

$$S_3(x_1, i) = \min\{x_2 \mid \pi_{3i} + V(x) - V(x - e_1) \geq 0, \text{ or}$$

$$\pi_{3i} + V(x) - V(x - e_2) \geq 0, \text{ given } x_1\},$$

$$S_4(x_1, i) = \min\{x_2 \mid V(x - e_1) - V(x - e_2) \geq 0, \text{ given } x_1\},$$

$$S_5(x_2, j) = \min\{x_1 \mid c_{1j} + V(x + e_1) - V(x) \geq 0, \text{ given } x_2\},$$

$$S_6(x_1, k) = \min\{x_2 \mid c_{2k} + V(x + e_2) - V(x) \geq 0, \text{ given } x_1\}.$$

From the above definitions, S_1 and S_2 are associated with the acceptance/rejection decisions for types 1 and 2 demand, respectively. S_3 and S_4 are associated with acceptance/rejection and routing controls for type 3 demands. The switching functions S_5 and S_6 are for production control of factories A and B, respectively. To ease notations, we occasionally suppress *subscripts* in functions $S(\cdot, \cdot)$ in the following discussion.

Lemma 4. *Functions $S_1(x_2)$ and $S_5(x_2)$ are decreasing in x_2 ; $S_2(x_1)$, $S_3(x_1)$ and $S_6(x_1)$ are decreasing in x_1 ; $S_4(x_1)$ is increasing in x_1 .*

Proof. Please see the appendix. □

Lemma 4 characterizes the structure of optimal policies for stock rationing and production control. For the single-factory problem, the optimal rationing policy is typically characterized as multiple critical levels. In contrast, the structure of the optimal policy for the two-factory problem is characterized by multiple monotone switching curves. Consequently, simply applying the multiple critical levels policy to the two-factory problem could yield a significant deviation from the optimal solution. This will be illustrated by numerical examples in Sec. 5.

Theorem 1. *The optimal policies are characterized by the monotone switching curves $S_1(x_2), S_2(x_1), S_3(x_1), S_4(x_1), S_5(x_2)$ and $S_6(x_1)$ according to the following decision rules, given the state (x_1, x_2) :*

- (i) *Type I demand of class i will be accepted if and only if $x_1 \geq S_1(x_2)$;*
- (ii) *Type II demand of class i will be accepted if and only if $x_2 \geq S_2(x_1)$;*
- (iii) *For Type III demand of class i , there are four cases:*
 - (a) *satisfy the demand with items of factory B if $x_2 \geq S_3(x_1)$ and $x_2 > S_4(x_1)$;*
 - (b) *satisfy the demand with items of factory A if $x_2 \geq S_3(x_1)$ and $x_2 < S_4(x_1)$;*
 - (c) *satisfy the demand with items of either factory A or B if $x_2 \geq S_3(x_1)$ and $x_2 = S_4(x_1)$;*
 - (d) *reject the demand if $x_2 < S_3(x_1)$.*
- (iv) *For the production at factory A, employ all the production rates $\mu_{11}, \mu_{12}, \dots, \mu_{1,j-1}, \mu_{1j}$ when $x_1 < S_5(x_2, j)$, for $j = 1, 2, \dots, n_1$.*
- (v) *For the production at factory B, employ all the production rates $\mu_{21}, \mu_{22}, \dots, \mu_{2,k-1}, \mu_{2k}$ when $x_2 < S_6(x_1, k)$, for $k = 1, 2, \dots, n_2$.*

Proof. Please see the appendix. □

Here, in Figs. 3 and 4, x_1 and x_2 denote the inventory levels of factories A and B, respectively. The solid curves divide the state space into several decision zones. In Fig. 4, the dotted lines are of lower production cost compared with the solid curves of relative higher production cost. This is the monotone property of $S_5(x_2, j)$ and $S_6(x_1, k)$ regarding the production rates and is proved in the following proposition.

Proposition 1. *$S_1(x_2, i), S_2(x_1, i)$ and $S_3(x_1, i)$ are increasing in demand classes; $S_5(x_2, j)$ and $S_6(x_1, k)$ are decreasing in production rates.*

Fig. 3. The switching curves for controlling three types of demand.

Fig. 4. The switching curves for the production controls.

Proof. Please see the appendix. $\square$

From Proposition 1, we know that the switching curves for each type of demand are nested in m classes and production switching curves are nested in the classes of discrete production rates. Intuitively, since $V(x+1)-V(x)$ is increasing in x_2 , with the fixed x_1 , it will surpass $-\pi_i$ one by one with respect to demand classes i as x_2 increases. It also implies that demands in lower classes are more likely to be rejected. Regarding the production process, production rate with the lowest cost is first used. Higher production rates with higher costs could be employed one by one when inventory level goes down across certain thresholds since the switching curves associated with production controls are monotone decreasing in the production rates.

The following proposition implies that it is always optimal to accept the demand of the highest priority if inventory is still available.

Proposition 2. (i) $V(x - e_1) \leq V(x) + \pi_{11}$ and $V(x - e_1) \leq V(x) + \pi_{31}$, if $x_1 > 0$;
(ii) $V(x - e_2) \leq V(x) + \pi_{21}$ and $V(x - e_2) \leq V(x) + \pi_{31}$, if $x_2 > 0$.

Proof. Please see the appendix. $\square$

3.2. The batch demand case

In the previous model formulation, we have assumed that the demand of each class and type follows the Poisson process and characterized the optimal policy as a switching-curve type. In this subsection, we extend the analysis from previous single-unit demand case to the batch demand case. That is, the demand is now described by the compound Poisson process. We still assume that incoming customers of three types follow the Poisson processes with arrival rates λ_{1i} , λ_{2i} and

λ_{3i} ($i = 1, \dots, m$), respectively. Moreover, each arrival is supposed to request a batch size a_{li} ($1 \leq a_{li} \leq M_{li}$) with the probability $p_{a_{li}}$, where M_{li} is the largest demand size for type l and class i and $\sum_{a_{li}=1}^{M_{li}} p_{a_{li}} = 1$, $l = 1, 2, 3$. We further assume that the batch demand can be partially satisfied, that is, given the demand size a_{li} units, we can satisfy any quantity between 0 and a_{li} if stock on hand is sufficient to meet the demand. When a customer arrives, her demand is either satisfied (fully or partially) or rejected, and rejected demand is lost. As before, we continue to assume that $\sum_{l=1}^3 \sum_{i=1}^m \lambda_{li} \sum_{a_{li}=1}^{M_{li}} p_{a_{li}} a_{li} < \sum_{j=1}^{n_1} \mu_{1j} + \sum_{k=1}^{n_2} \mu_{2k}$, that is, the total production capacity can handle the total demand in the long run. Since we only alter the demand size in this case, the uniformized transition rate $\Lambda = \sum_{i=1}^m (\lambda_{1i} + \lambda_{2i} + \lambda_{3i}) + \sum_{j=1}^{n_1} \mu_{1j} + \sum_{k=1}^{n_2} \mu_{2k}$ for the underlying demand arrival and production processes remains unchanged.

All the other notations and model formulation are the same with the single-unit demand case except for the operator T . Thus, (1) is reformulated as follows:

$$\begin{aligned} TV = & h + \sum_{i=1}^m \lambda_{1i} \sum_{a_{1i}=1}^{M_{1i}} p_{a_{1i}} T_{1,i}^{a_{1i}} V + \sum_{i=1}^m \lambda_{2i} \sum_{a_{2i}=1}^{M_{2i}} p_{a_{2i}} T_{2,i}^{a_{2i}} V \\ & + \sum_{i=1}^m \lambda_{3i} \sum_{a_{3i}=1}^{M_{3i}} p_{a_{3i}} T_{3,i}^{a_{3i}} V + \sum_{j=1}^{n_1} \mu_{1j} T_{4,j} V + \sum_{k=1}^{n_2} \mu_{2k} T_{5,k} V, \end{aligned}$$

where $T_{l,i}^{a_{li}}$, $l = 1, 2, 3$, indicates the composition of $T_{l,i}$ with itself a_{li} times and $T_{l,i}$, $l = 1, 2, 3$ and $i = 1, \dots, m$, are the operators defined in Sec. 2.

In the following, we illustrate the derivation process of the similar results in parallel to those in Theorem 1. First, in an analogous approach to Lemmas 1–2, we can prove that T retains the structural properties (a), (b) and (c) and $V \in \hat{V}$. Then, as before, we can define the switching functions for rationing controls of production capacities and inventories. Next, with the same argument as that in the proof of Lemma 4 we can show that all the switching functions are monotone. Finally, we can again characterize the optimal joint stock and production capacity rationing policies as monotone switching curves. Since the proofs are similar to those in subsection 3.1, we state the main results in the following Theorem 2 without a proof.

Theorem 2. *The optimal policies are characterized by $S_1(x_2), S_2(x_1), S_3(x_1), S_4(x_1), S_5(x_2)$ and $S_6(x_1)$ with the following decision rules: at a given state (x_1, x_2) .*

- (i) *For the demand size a_{1i} ($1 \leq a_{1i} \leq M_{1i}$) of Type I and class i , the demand will be accepted if and only if $x_1 \geq S_1(x_2)$; if it is accepted, then applying the same rule to the state $(x_1 - 1, x_2)$, and this rule is applied repeatedly until the demand is rejected or the full size a_{1i} is accepted.*
- (ii) *For the demand size a_{2i} ($1 \leq a_{2i} \leq M_{2i}$) of Type II and class i , the demand will be accepted if and only if $x_2 \geq S_2(x_1)$; if it is accepted, then applying the*

same rule to the state $(x_1, x_2 - 1)$, and this rule is applied repeatedly until the demand is rejected or the full size a_{2i} is accepted.

- (iii) For the demand size a_{3i} ($1 \leq a_{3i} \leq M_{3i}$) of Type III and class i , there are the following cases:
- (a) Satisfy the demand with items of factory B if $x_2 \geq S_3(x_1)$ and $x_2 > S_4(x_1)$ and change the state from (x_1, x_2) to $(x_1, x_2 - 1)$.
 - (b) Satisfy the demand with items of factory A if $x_2 \geq S_3(x_1)$ and $x_2 < S_4(x_1)$ and change the state from (x_1, x_2) to $(x_1 - 1, x_2)$.
 - (c) Satisfy the demand with items of either factory A or B if $x_2 \geq S_3(x_1)$ and $x_2 = S_4(x_1)$ and correspondingly change the state from (x_1, x_2) either to $(x_1 - 1, x_2)$ or to $(x_1, x_2 - 1)$.
 - (d) Reject the demand if $x_2 < S_3(x_1)$ and stop; otherwise, apply the steps (a), (b) and (c) repeatedly until the full-size demand a_{3i} is accepted.
- (iv) For the production at factory A, employ all the production rates $\mu_{11}, \mu_{12}, \dots, \mu_{1j-1}, \mu_{1j}$ when $x_1 < S_5(x_2, j)$, for $j = 1, 2, \dots, n_1$.
- (v) For the production at factory B, employ all the production rates $\mu_{21}, \mu_{22}, \dots, \mu_{2k-1}, \mu_{2k}$ when $x_2 < S_6(x_1, k)$, for $k = 1, 2, \dots, n_2$.

4. Analysis of the Flexible Demand

We investigate the value of flexible demand by a comparison. For the case without flexible demand, Type III customers are separated into Type I and Type II. Without loss of generality, assume $\lambda_{3i} = \gamma_{1i} + \gamma_{2i}$, $i = 1, 2, \dots, m$, where

γ_{1i} = the class i arrival rate of the original type III demand assigned for factory A,
 γ_{2i} = the class i arrival rate of the original type III demand assigned for factory B.

For this problem, the uniformized transition rate becomes

$$\Gamma = \sum_{i=1}^m (\lambda_{1i} + \lambda_{2i} + \gamma_{1i} + \gamma_{2i}) + \sum_{j=1}^{n_1} \mu_{1j} + \sum_{k=1}^{n_2} \mu_{2k}.$$

Obviously, Γ equals Λ in Sec. 2. Denote the optimal cost function by $U(x)$. Let $\alpha + \Gamma = 1$, then the Bellman equation is

$$U = T'U, \tag{3}$$

where T' is defined by

$$\begin{aligned} T'U = & h + \sum_{i=1}^m (\lambda_{1i} + \gamma_{1i})T_{1,i}U + \sum_{i=1}^m (\lambda_{2i} + \gamma_{2i})T_{2,i}U \\ & + \sum_{j=1}^{n_1} \mu_{1j}T_{4,j}U + \sum_{k=1}^{n_2} \mu_{2k}T_{5,k}U, \end{aligned} \tag{4}$$

where $T_{1,i}, T_{2,i}, T_{4,j}, T_{5,k}$ have the same definitions as those in Sec. 2. Under certain condition, we show the optimal cost associated with the flexible demand is no greater than U .

Proposition 3. *If $\pi_{3i} \leq \min\{\pi_{1i}, \pi_{2i}\}$, (1) $V(x) \leq U(x)$ for all x ; (2) $V(x) \downarrow \lambda_{3i}$ for fixed $\bar{\lambda}$.*

Proof. Please see the appendix. $\square$

The first result (1) of Proposition 3 states that if penalty costs for denied flexible demand are no greater than those for denied demands of Types I and II, segmentation of flexible demand can reduce the optimal cost. The second result implies that a higher ratio of flexible demand yields a lower optimal cost. Thus, it is beneficial for firms to have or induce more customers to be flexible about their products.

Furthermore, the splitting of flexible demand makes Types I and II demands as well as productions at two factories to be independent. As a result, we can develop a heuristic approach by decomposing the original problem into two independent ones, one for Type I demand and factory A, and the other for Type II and factory B. As before we rescale the time to achieve $\alpha + \Gamma = 1$ and the corresponding Bellman equations are $U_A = T_A U_A$ and $U_B = T_B U_B$. The operators T_A and T_B are defined by

$$\begin{aligned} T_A U_A = & h_1 + \sum_{i=1}^m (\lambda_{1i} + \gamma_{1i}) T_{1,i}^A U_A + \sum_{j=1}^{n_1} \mu_{1j} T_{4,j}^A U_A \\ & + \left[\sum_{i=1}^m (\lambda_{2i} + \gamma_{2i}) + \sum_{k=1}^{n_2} \mu_{2k} \right] U_A, \end{aligned} \quad (5)$$

where

$$T_{1,i}^A U_A(x_1) = \begin{cases} \min\{U_A(x_1 - 1), U_A(x_1) + \pi_{1i}\} & \text{if } x_1 > 0 \\ U_A(x_1) + \pi_{1i} & x_1 = 0 \end{cases},$$

$$T_{4,j}^A U_A(x_1) = \min\{c_{1j} + U_A(x_1 + 1), U_A(x_1)\}.$$

$$\begin{aligned} T_B U_B = & h_2 + \sum_{i=1}^m (\lambda_{2i} + \gamma_{2i}) T_{2,i}^B U_B + \sum_{k=1}^{n_2} \mu_{2k} T_{5,k}^B U_B \\ & + \left[\sum_{i=1}^m (\lambda_{1i} + \gamma_{1i}) + \sum_{j=1}^{n_1} \mu_{1j} \right] U_B, \end{aligned} \quad (6)$$

where

$$T_{2,i}^B U_B(x_2) = \begin{cases} \min\{U_B(x_2 - 1), U_B(x_2) + \pi_{2i}\} & \text{if } x_2 > 0 \\ U_B(x_2) + \pi_{2i} & x_2 = 0 \end{cases},$$

$$T_{5,k}^B U_B(x_2) = \min\{c_{2k} + U_B(x_2 + 1), U_B(x_2)\}.$$

We can characterize the optimal policy for each of the above two problems as critical levels, which are simpler in structures and easier for implementation than

the switching-curve policy. Moreover, we show the following result that is the basis for the decomposition method.

Proposition 4. $U(x_1, x_2) = U_A(x_1) + U_B(x_2)$.

Proof. The result is easily proved by the value iteration method. $\square$

5. Numerical Analysis

Without loss of generality, we consider the case of two classes for each type of demand: high priority and low priority classes with $\pi_{11} = 30, \pi_{12} = 15, \pi_{21} = 30, \pi_{22} = 15, \pi_{31} = 30$ and $\pi_{32} = 15$; and two production modes: low cost production and high cost production with the unit production costs $c_{11} = 10, c_{12} = 15, c_{21} = 10$ and $c_{22} = 15$. The unit holding cost rates are assumed to be $h_1 = 1, h_2 = 1$. Let the discount rate $\alpha = 0.1$. The arrival rates and production rates of different types and classes are listed in Table 1.

By the value iterations, we compute the optimal costs and policies starting with $V_0(x) = 0$ and stopping at $V_{980}(0, 0) = 1326.614, V_{981}(0, 0) = 1326.657$, with $(V_{981}(0, 0) - V_{980}(0, 0))/V_{980}(0, 0) = 0.003293\%$. Some of the optimal decisions restricted to the states of $0 \leq x_1 \leq 15$ and $0 \leq x_2 \leq 15$ are presented in Tables 2 and 3.

In Table 2, we use 0, 1, 2 to denote the rejection, satisfying the demand by factory A and satisfying the demand by factory B, respectively. The above two figures show the switching curves for low priority classes of Types I and II demands. The switching curves for high priority class are not presented here. But from Proposition 2, it is conceivable that they have the form of straight lines with $x_1 = 1$ or $x_2 = 1$, which implies it is always optimal to accept the high priority class demand if stocks are still available. For the flexible demand, the monotone decreasing curve $S_3(x_1, 2)$ combined with the monotone increasing curve $S_4(x_1, 2)$ divide the state space into three decision regions that are used for the admission/routing of flexible demand. If current inventory state is below $S_3(x_1, 2)$ which is the area with decision 0 in the table, it is optimal to deny the flexible demand. Otherwise, accept the flexible demand. Upon acceptance of the flexible demand, we use $S_4(x_1, 2)$ to dispose it. If the inventory state is in the region with decision 2, which is just above $S_4(x_1, 2)$, it is optimal to satisfy the demand by factory B for the reason that inventory of

Table 1. The arrival rates and production rates.

Arrival rates of	Type 1 demand	Type 2 demand	Type 3 demand
High Priority	2.5	2	2
Low Priority	1	1	1
Production rates of	Factory A	Factory B	
Low cost	3	3	
High cost	2	2	

Table 2. The decisions for the arrival processes.

factory B is relatively higher than that of factory A. When the inventory state is in the region with decision 1, it is optimal to satisfy the demand by factory A, which has relatively higher inventory than that of factory B.

For the production processes, Table 3 illustrates the idleness/ production decisions at factory A and factory B, denoted by 0/1 and 0/2, respectively.

In Table 3, the upper two figures shows the production switching curves for factory A with the lower cost (the left curve) and the higher cost (the right curve), respectively. The rest two figures show the switching curves of lower cost and higher cost for factory B. It is evident that production curves are monotone decreasing with respect to the production cost. Intuitively, management can take overtime production or more production capacities in terms of subcontracting when inventory on hand runs relatively low.

Next, we study the value of flexible demand based on sensitivity analysis of holding cost rates and arrival rates (with total arrival rates $\bar{\lambda}$ fixed and other cost parameters unchanged). In Table 4, note that λ_{31} and λ_{32} refer to the arrival rates of the high priority and low priority flexible demands, respectively.

For the cases 1–6, we split the flexible demand into half with $\gamma_{1i} = \gamma_{2i} = 0.5\lambda_{3i}$. Case 1 shows a 3.5% increase of optimal cost compared with the base case.

Table 3. The decisions for the production processes.

the switching curve for the low cost production of factory A																
x_2	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
15	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
14	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
13	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
12	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
11	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
10	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
9	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
8	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
7	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
6	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
5	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
4	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
3	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
2	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
x_1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

the switching curve for the high cost production of factory A																
x_2	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
15	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
x_1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

the switching curve for the low cost production of factory B																
x_2	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0
8	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0
7	2	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0
6	2	2	2	2	2	2	2	2	0	0	0	0	0	0	0	0
5	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0
4	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
x_1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

the switching curve for the high cost production of factory B																
x_2	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0
1	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0
0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
x_1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Table 4. The value of flexible demand.

	Holding cost rates and arrival rates												Comparison	
	h_1	h_2	λ_{11}	λ_{12}	λ_{21}	λ_{22}	λ_{31}	λ_{32}	μ_{11}	μ_{12}	μ_{21}	μ_{22}	$V(0,0)$	Variation (%)
Base case	1	1	2.5	1	2	1	2	1	3	2	3	2	1326.614	0
1	1	1	3.5	1.5	3	1.5	0	0	3	2	3	2	1373.059	3.5
2	0.5	0.5	3.5	1.5	3	1.5	0	0	3	2	3	2	1342.843	1.22
3	0.25	0.25	3.5	1.5	3	1.5	0	0	3	2	3	2	1325.821	-0.06
4	0.1	0.1	3.5	1.5	3	1.5	0	0	3	2	3	2	1315.515	-0.84
5	2	2	3.5	1.5	3	1.5	0	0	3	2	3	2	1425.447	7.45
6	5	5	3.5	1.5	3	1.5	0	0	3	2	3	2	1552.813	17.05
7	1	1	2	1	2	1	2.5	1	3	2	3	2	1318.751	-0.59
8	1	1	1.5	1	1.5	1	3.5	1	3	2	3	2	1305.812	-1.57
9	1	1	1	1	1	1	4.5	1	3	2	3	2	1294.461	-2.42
10	1	1	1	0.5	1	0.5	4.5	2	3	2	3	2	1293.905	-2.47
11	1	1	0.5	0.5	0.5	0.5	5.5	2	3	2	3	2	1284.955	-3.14
12	1	1	2.5	1	2	1	2	1	4	1	4	1	1254.43	-5.44
13	1	1	2.5	1	2	1	2	1	5	2	3	2	1238.877	-6.61
14	1	1	2.5	1	2	1	2	1	2	3	2	3	1415.087	6.67
15	1	1	2.5	1	2	1	2	1	2.4	2.4	2.4	2.4	1384.292	4.35
16	1	1	2.5	1	2	1	2	1	5	2	5	2	1177.161	-11.27

From the cases 1–6, we see that the variations of optimal cost are sensitive to the increment of h_1 and h_2 but insensitive to the decreasing of h_1 and h_2 . For the cases 7–11, we increase the ratio $\lambda_{3i}/\bar{\lambda}$, leading to a reduction of optimal cost. Thus, an increasing flexible demand within the total demand implies a decreasing optimal cost. For the cases 12–16, we apply sensitivity analysis to production rates. Generally, if we increase the low cost production rates the optimal cost decreases correspondingly; if we increase the high cost production rates the optimal cost increases correspondingly.

As discussed in Sec. 4, we can develop a heuristic approach to the original problem by classifying the flexible demand into either Type I or Type II demand, which leads to two single-factory problems. For each single-factory problem, we can characterize the optimal policy with critical levels that help to reduce computation workload. Correspondingly, the optimal cost $U(x_1, x_2)$ for the decomposed problem is the sum of the optimal costs $U_A(x_1)$ and $U_B(x_2)$ for two single-factory problems, i.e., $U(x_1, x_2) = U_A(x_1) + U_B(x_2)$. In the sequel we shall divide the flexible demand by the so-called Bernoulli-splitting rule. Under this rule each arrival is assigned with a given probability of being sent to a factory. Let β be the probability of a flexible demand classified into Type I to factory A and thus $1 - \beta$ is the probability of flexible demand classified into Type II to factory B. As a consequence, the arrival

Table 5. Comparison of the optimal and heuristic policies.

	β	Holding cost rates and arrival rates												Comparison	
		h_1	h_2	λ_{11}	λ_{12}	λ_{21}	λ_{22}	λ_{31}	λ_{32}	μ_{11}	μ_{12}	μ_{21}	μ_{22}	$V(0,0) \& U(0,0)$	Variation (%)
Base case 1		1	1	2.5	1	2	1	2	1	3	2	3	2	1326.614	0
1	0.3	1	1	3.1	1.3	3.4	1.7	0	0	3	2	3	2	1372.575	3.46
2	0.5	1	1	3.5	1.5	3	1.5	0	0	3	2	3	2	1373.059	3.5
3	0.7	1	1	3.9	1.7	2.6	1.3	0	0	3	2	3	2	1380.198	4.04
Base case 2		0.5	0.5	2.5	1	2	1	2	1	3	2	3	2	1300.51	0
4	0.3	0.5	0.5	3.1	1.3	3.4	1.7	0	0	3	2	3	2	1342.168	3.2
5	0.5	0.5	0.5	3.5	1.5	3	1.5	0	0	3	2	3	2	1342.843	3.26
6	0.7	0.5	0.5	3.9	1.7	2.6	1.3	0	0	3	2	3	2	1350.44	3.84
Base case 3		2	2	2.5	1	2	1	2	1	3	2	3	2	1373.79	0
7	0.3	2	2	3.1	1.3	3.4	1.7	0	0	3	2	3	2	1425.846	3.79
8	0.5	2	2	3.5	1.5	3	1.5	0	0	3	2	3	2	1425.446	3.76
9	0.7	2	2	3.9	1.7	2.6	1.3	0	0	3	2	3	2	1433.276	4.33
Base case 4		1	1	1	1	1	1	4.5	1	3	2	3	2	1294.461	0
10	0.1	1	1	1.45	1.1	5.05	1.9	0	0	3	2	3	2	1438.907	11.16
11	0.3	1	1	2.35	1.3	4.15	1.7	0	0	3	2	3	2	1388.345	7.25
12	0.5	1	1	3.25	1.5	3.25	1.5	0	0	3	2	3	2	1371.029	5.92
Base case 5		1	1	0.5	0.5	0.5	0.5	5.5	2	3	2	3	2	1284.955	0
13	0.1	1	1	1.05	0.7	5.45	2.3	0	0	3	2	3	2	1482.526	15.4
14	0.3	1	1	2.15	1.1	4.35	1.9	0	0	3	2	3	2	1401.014	9.03
15	0.5	1	1	3.25	1.5	3.25	1.5	0	0	3	2	3	2	1371.029	6.7

rates of Type I and Type II demands become $\sum_{i=1}^m (\lambda_{1i} + \beta\lambda_{3i})$ and $\sum_{i=1}^m \{\lambda_{2i} + (1 - \beta)\lambda_{3i}\}$, respectively. For the convenience of sensitivity analysis on the arrival rates in the following proposed cases, we apply the finite-horizon dynamic programming to compute the optimal cost functions given the state at the original point $(0, 0)$.

Table 5 shows the comparison results of the optimal and heuristic policies. In group 1 (with base case 1), group 2 (with base case 2) and group 3 (with base case 3), unit holding costs are different. It is obvious that the variations of the optimal costs are increasing with regard to unit holding costs no matter how we divide the flexible demands. In groups 4 and 5, the flexible demands account for a large portion of the total demands. Compared with those of group 1, 2 and 3, variations of the optimal costs in groups 4 and 5 are significantly increasing no matter how we divide the flexible demand. Another observation is that $\beta = 0.5$ is a better choice for nearly identical parameters at two factories. Moreover, if $\pi_{3i} \leq \min\{\pi_{1i}, \pi_{2i}\}$, from the result (1) of Proposition 3, we can calculate the optimal cost for different selected β and then choose the smallest optimal cost to be the best approximation to that of the original problem. And the computation workload grows linearly with the number of selected β value.

6. Conclusion

In this paper, we study both stock and production capacity rationings in a make-to-stock system with two factories and two products. Based on the structure properties of the optimal cost function, we characterize the structure of the optimal policy as switching curves, which separate the state space into several decision zones. Monotone properties for the switching curves have also been derived. An interesting attribute of our problem is the introduction of flexible demand, which is indifferent between the items manufactured by two factories. Moreover, it has been proved that the segmentation of flexible demand benefits in reducing the total costs. Essentially, rationing method is to differentiate demands according to their economic value to the suppliers. The classification of the flexible demand is in vein of the rationing method.

In this paper, we only characterize the optimal joint stock and capacity rationing policies for the lost-sales case, which limits the practical application of our model to some extent. Further research on the backlog case of two dedicated types of demand is more practical and challenging. Another interesting research direction is to further consider shipping costs in addition to the production and inventory costs in the two-factory problem with identical products. As such, we shall allow three types of demand to be fulfilled by products either from plant A or from B in order to achieve a more effective cost reduction. Although we present only the analysis for the two-factory case, our model and analysis are conveniently extended to the case with more than two factories which can be addressed by the same decomposition method proposed in this paper.

Appendix

Proof of Lemma 2. First, we prove $\Delta_1 V(x) \uparrow x_1$ in (a), i.e., $V(x + e_1) - V(x) \leq V(x + 2e_1) - V(x + e_1)$.

Let $V_0 = 0$. Evidently, $V_0(x + e_1) - V_0(x) \leq V_0(x + 2e_1) - V_0(x + e_1)$. Then we can obtain from Lemma 1 that $TV_0(x)$ preserve the inequality, i.e.,

$$TV_0(x + e_1) - TV_0(x) \leq TV_0(x + 2e_1) - TV_0(x + e_1).$$

Applying T repeatedly, we have by induction

$$T^n V_0(x + e_1) - T^n V_0(x) \leq T^n V_0(x + 2e_1) - T^n V_0(x + e_1).$$

Since the state space is discrete, for each fixed x , $T^n V_0(x) \rightarrow V(x)$ when $n \rightarrow \infty$. Therefore, $T^n V_0$ is point-wise convergent to V for all x . As a result, for all x , $V(x + e_1) - V(x) \leq V(x + 2e_1) - V(x + e_1)$, i.e., $\Delta_1 V(x) \uparrow x_1$. The rest structural properties can be verified analogously. Hence, $V \in \hat{V}$. $\square$

Proof of Lemma 3. Since the action set is finite, we can find an admissible policy u to achieve the minimum of the right-hand side of (2). As the cost per stage is non-negative, according to Bertsekas (1995), u is optimal. $\square$

Proof of Lemma 4. According to the definition of $S_1(x_2, i)$, given any fixed x_2 , $S_1(x_2, i)$ is the least value of x_1 to satisfy the following inequality

$$\pi_{1i} + V(x_1, x_2) - V(x_1 - 1, x_2) \geq 0. \quad (\text{A.1})$$

Therefore,

$$\pi_{1i} + V(S_1(x_2, i), x_2) - V(S_1(x_2, i) - 1, x_2) \geq 0. \quad (\text{A.2})$$

Based on $\Delta_1 V(x) \uparrow x_2$ in (a), i.e., supermodularity of $V(x)$, we have

$$V(x_1, x_2) - V(x_1 - 1, x_2) \leq V(x_1, x_2 + 1) - V(x_1 - 1, x_2 + 1). \quad (\text{A.3})$$

Hence, based on (A.2) and (A.3),

$$\begin{aligned} & \pi_{1i} + V(S_1(x_2, i), x_2) - V(S_1(x_2, i) - 1, x_2) \\ & \leq \pi_{1i} + V(S_1(x_2, i), x_2 + 1) - V(S_1(x_2, i) - 1, x_2 + 1) \end{aligned}$$

and

$$\pi_{1i} + V(S_1(x_2, i), x_2 + 1) - V(S_1(x_2, i) - 1, x_2 + 1) \geq 0. \quad (\text{A.4})$$

On the other hand, according to the definition of switching function $S_1(x_2, i)$, given a fixed $(x_2 + 1)$, $S_1(x_2 + 1, i)$ is the least value of x_1 to satisfy

$$\pi_{1i} + V(S_1(x_2 + 1, i), x_2 + 1) - V(S_1(x_2 + 1, i) - 1, x_2 + 1) \geq 0. \quad (\text{A.5})$$

After comparing the first argument of $V(x_1, x_2 + 1)$ in (A.4) and (A.5), we get $S_1(x_2, i) \geq S_1(x_2 + 1, i)$ based on the fact that $S_1(x_2 + 1, i)$ is the least value of x_1

to satisfy

$$\pi_{1i} + V(x_1, x_2 + 1) - V(x_1 - 1, x_2 + 1) \geq 0.$$

Since x_2 is arbitrarily given, $S_1(x_2, i) \geq S_1(x_2 + 1, i)$ implies that $S_1(x_2, i)$ is decreasing in x_2 . Analogously, we can prove that $S_5(x_2)$ is decreasing in x_2 and $S_2(x_1)$, $S_3(x_1)$ and $S_6(x_1)$ decreasing in x_1 .

Finally, we prove the result for $S_4(x_1)$. For a given fixed x_1 , by the definition, $S_4(x_1)$ is the least value of x_2 to satisfy

$$V(x_1 - 1, S_4(x_1)) - V(x_1, S_4(x_1) - 1) \geq 0. \quad (\text{A.6})$$

From the property $\Delta_1 V(x) - \Delta_2 V(x) \uparrow x_1$ in (c), we can obtain

$$V(x_1 - 1, x_2 - 1) - V(x_1 - 2, x_2) \leq V(x_1, x_2 - 1) - V(x_1 - 1, x_2). \quad (\text{A.7})$$

After rearranging the terms of (A.7),

$$V(x_1 - 1, x_2) - V(x_1, x_2 - 1) \leq V(x_1 - 2, x_2) - V(x_1 - 1, x_2 - 1). \quad (\text{A.8})$$

Hence, from (A.7) and (A.8),

$$\begin{aligned} V(x_1 - 2, S_4(x_1)) - V(x_1 - 1, S_4(x_1) - 1) \\ \geq V(x_1 - 1, S_4(x_1)) - V(x_1, S_4(x_1) - 1) \end{aligned}$$

and

$$V(x_1 - 2, S_4(x_1)) - V(x_1 - 1, S_4(x_1) - 1) \geq 0. \quad (\text{A.9})$$

On the other hand, from the definition of $S_4(x_1)$, for the fixed $(x_1 - 1)$, $S_4(x_1 - 1)$ is the least value of x_2 to satisfy

$$V(x_1 - 2, S_4(x_1 - 1)) - V(x_1 - 1, S_4(x_1 - 1) - 1) \geq 0. \quad (\text{A.10})$$

After comparing the second argument of $V(x_1 - 2, x_2)$ in (A.9) with (A.10) and noting the fact that $S_4(x_1 - 1)$ is the least value to satisfy

$$V(x_1 - 2, x_2) - V(x_1 - 1, x_2 - 1) \geq 0.$$

Hence, $S_4(x_1) \geq S_4(x_1 - 1)$ for an arbitrarily given x_1 , i.e., $S_4(x_1)$ is increasing in the argument x_1 . $\square$

Proof of Theorem 1. For (i), from Lemma 4, $S_1(x_2)$ is decreasing in x_2 . Based on the convexity of the optimal cost function and the definition of $S_1(x_2)$, we know that $\pi_{1i} + V(x) - V(x - e_1) \geq 0$ whenever $x_1 \geq S_1(x_2)$. Thus, it is optimal to satisfy Type I demand whenever $x_1 \geq S_1(x_2)$ and reject the demand otherwise.

Similarly, (ii) can be proved.

For (iii), from Lemma 4, $S_3(x_1)$ is decreasing in x_1 . Above $S_3(x_1)$, for fixed x_1 , we have either $\pi_{3i} + V(x) - V(x - e_1) \geq 0$ or $\pi_{3i} + V(x) - V(x - e_2) \geq 0$. Hence, it is optimal to accept Type III demand, i.e., the flexible demand. Below $S_3(x_1)$, it is optimal to reject it. $S_4(x_1)$ is used for disposing the accepted Type III demand.

On the above of $S_4(x_1)$, for any fixed x_1 , we have $V(x - e_1) - V(x - e_2) \geq 0$. This is true from structural property (c). Therefore, it is optimal to satisfy the flexible demand with items of factory B. Below $S_4(x_1)$, it is optimal to satisfy the demand with items of factory A. The switching curves are illustrated in Fig. 3.

For (iv), on the right-hand side of $S_5(x_2)$, for fixed $x_2, x_1 \geq S_5(x_2)$. According to the definition of $S_5(x_2)$ and the property (a), we have $c_{1j} + V(x + e_1) - V(x) \geq 0$ on the right side of the curve (i.e., $x_1 \geq S_5(x_2)$). Hence, it is optimal to stop the production. On the left side of $S_5(x_2)$, i.e., $x_1 < S_5(x_2)$, we have $c_{1j} + V(x + e_1) - V(x) < 0$. It is also obvious $c_{1n} + V(x + e_1) - V(x) < 0$, for $n = 1, 2, \dots, j - 1$, that is, the inequality is also true for production costs lower than c_{1j} . Hence, it is optimal to produce item 1 with all the production rates $\mu_{11}, \mu_{12}, \dots, \mu_{1j-1}, \mu_{1j}$ when $x_1 < S_5(x_2)$.

Similarly, (v) can be proved. The switching curves are illustrated in Fig. 4. $\square$

Proof of Proposition 1. First, consider $S_1(x_2, i) = \min\{x_1 \mid \pi_{1i} + V(x) - V(x - e_1) \geq 0, \text{ given } x_2\}$. Since $\pi_{1i} > \pi_{1,i+1}$, we have $\pi_{1i} + V(S_1(x_2, i + 1), x_2) - V(S_1(x_2, i + 1) - 1, x_2) > 0$. Then the result $S_1(x_2, i) \leq S_1(x_2, i + 1)$ is immediate. Analogously, results can be derived for $S_2(x_1, i), S_3(x_1, i), S_5(x_2, j)$ and $S_6(x_1, k)$. $\square$

Proof of Proposition 2. We only need to prove (i) while (ii) can be derived analogously. We begin with $V(x - e_1) \leq V(x) + \pi_{11}$ in (i). First, it needs to show $TV(x - e_1) - TV(x) \leq \pi_{11}$, i.e., the inequality is preserved by the operator T .

First, we show $T_{1,i}$ preserves $V(x - e_1) \leq V(x) + \pi_{11}$, i.e., $T_{1,i}V(x - e_1) - T_{1,i}V(x) \leq \pi_{11}$, that is, $\min\{V(x - 2e_1), V(x - e_1) + \pi_{1i}\} - \min\{V(x - e_1), V(x) + \pi_{1i}\} \leq \pi_{11}$.

Since $V(x - e_1) + \pi_{1i} - V(x - e_1) \leq \pi_{11}$ and $V(x - e_1) + \pi_{1i} - V(x) - \pi_{1i} \leq \pi_{11}$, the result applies. Similarly, we can prove the results for $T_{2,i}, T_{3,i}, T_{4,j}$, and $T_{5,k}$. It is evident $h(x - e_1) - h(x) \leq \pi_{11}$. Thus, T preserves the inequality.

Second, we prove the result by value iteration. Let $V_0 = 0$. Obviously, the inequality is true for V_0 . Applying T repeatedly on V_0 , we derive that the inequality is true for $T^n V_0$ for all n . Letting $n \rightarrow \infty, T^n V_0(x) \rightarrow V(x)$ in point-wise convergence for all x . Hence, the inequality applies when $x_1 > 0$. Last, $V(x - e_1) \leq V(x) + \pi_{31}$ in (i) can be proved analogously. $\square$

Proof of Proposition 3. For (1), first, we prove $TV \leq T'U$. Noting that if $V(x) \leq U(x)$,

$$\min\{V(x - e_1), V(x - e_2), V(x) + \pi_{3i}\} \leq \min\{U(x - e_1), U(x) + \pi_{1i}\},$$

$$\min\{V(x - e_1), V(x - e_2), V(x) + \pi_{3i}\} \leq \min\{U(x - e_2), U(x) + \pi_{2i}\}.$$

Thus, $T_{3,i}V \leq T_{1,i}U$ and $T_{3,i}V \leq T_{2,i}U$. Moreover, it is convenient to show $T_{1,i}V \leq T_{1,i}U$, $T_{2,i}V \leq T_{2,i}U$, $T_{4,j}V \leq T_{4,j}U$, and $T_{5,k}V \leq T_{5,k}U$. Comparing the terms on

the right-hand side (RHS) of (1) and (3), we have $TV \leq T'U$. Furthermore, by the value iteration method, we obtain that $T_{1,i}^n V \leq T_{1,i}^n U$, $T_{2,i}^n V \leq T_{2,i}^n U$, $T_{3,i}^n V \leq T_{1,i}^n U$, $T_{3,i}^n V \leq T_{2,i}^n U$, $T_{4,j}^n V \leq T_{4,j}^n U$, and $T_{5,k}^n V \leq T_{5,k}^n U$ for all n . Since the uniformized transition rate Γ equals Λ ,

$$\begin{aligned} & (T')^n U - T^n V \\ &= \sum_{i=1}^m \lambda_{1i} (T_{1,i}^n U - T_{1,i}^n V) + \sum_{i=1}^m \lambda_{2i} (T_{2,i}^n U - T_{2,i}^n V) + \sum_{i=1}^m \gamma_{1i} (T_{1,i}^n U - T_{3,i}^n V) \\ & \quad + \sum_{i=1}^m \gamma_{2i} (T_{2,i}^n U - T_{3,i}^n V) + \sum_{j=1}^{n_1} \mu_{1j} (T_{4,j}^n U - T_{4,j}^n V) + \sum_{k=1}^{n_2} \mu_{2k} (T_{5,k}^n U - T_{5,k}^n V). \end{aligned}$$

Thus, $T^n V \leq (T')^n U$ for all n .

For any given x , $T^n V$ and $(T')^n U$ are the point-wise convergence to V and U respectively. Letting $n \rightarrow \infty$, the result applies.

For (2), noting that $\min\{V(x - e_1), V(x - e_2), V(x) + \pi_{3i}\} \leq \min\{V(x - e_1), V(x) + \pi_{1i}\}$ and $\min\{V(x - e_1), V(x - e_2), V(x) + \pi_{3i}\} \leq \min\{V(x - e_2), V(x) + \pi_{2i}\}$, we can prove the result in the same approach as that in (1). $\square$

Acknowledgments

The authors are grateful to the editor and two referees for their constructive comments and suggestions. The first author is supported by the scholar research fund of Jiangsu Normal University.

References

- Arslan, H, SC Graves and TA Roemer (2007). A single-product inventory model for multiple demand classes. *Management Science*, 53, 1486–1500.
- Axsäter, S, M Kleijn and TG de Kok (2004). Stock rationing in a continuous review two-echelon inventory model. *Annals of Operations Research*, 126, 177–194.
- Benjaafar, S and M ElHafsi (2006). Production and inventory control of a single product assemble-to-order system with multiple customer classes. *Management Science*, 52, 1896–1912.
- Bertsekas, DP (1995). *Dynamic Programming and Optimal Control*. Vol. II. Massachusetts: Athena Scientific.
- Bulut, Ö and MM Fadiloglu (2011). Production control and stock rationing for a make-to-stock system with parallel production channels. *IIE Transactions*, 43(6), 432–450.
- Cohen, MA, PR Kleindorfer and HL Lee (1988). Service constrained (s, S) inventory systems with priority demand classes and lost sales. *Management Science*, 34, 482–499.
- Chen, S (2004). The optimality of hedging point policies for stochastic two-product flexible manufacturing systems. *Operations Research*, 52, 312–322.
- Chen, S, J Xu and Y Feng (2010). A partial characterization of the optimal ordering/rationing policy for a periodic review system with two demand classes and back-ordering. *Naval Research Logistics*, 57(4), 330–341.

- Chew, EP, L Lee and S Liu (2013). Dynamic rationing and ordering policies for multiple demand classes. *OR Spectrum*, 35, 127–151.
- de Vericourt, F, FF Karaesmen and Y Dallery (2002). Optimal stock allocation for capacitated supply system. *Management Science*, 48, 1486–1500.
- Deshpande, V, MA Cohen and K Donohue (2003). A threshold inventory rationing policy for service-differentiated demand classes. *Management Science*, 49, 683–703.
- Donohue, K (2000). Efficient supply contracts for fashion goods with forecast updating and two production modes. *Management Science*, 46, 1397–1411.
- Enders, P, I Adan, A Scheller-Wolf and GJV Houtum (2014). Inventory rationing for a system with heterogeneous customer classes. *Flexible Services and Manufacturing Journal*, 26(3), 344–386.
- Fadiloglu, MM and Ö Bulut (2010). A dynamic rationing policy for continuous-review inventory systems. *European Journal of Operational Research*, 202(3), 675–685.
- Frank, KC, RC Zhang and I Duenyas (2003). Optimal inventory policy for inventory system with priority demand classes. *Operations Research*, 51, 993–1002.
- Goyal, M and S Netessine (2011). Volume flexibility, product flexibility, or both: The role of demand correlation and product substitution. *Manufacturing & Service Operations Management*, 13(2), 180–193.
- Ha, AY (1997a). Inventory rationing in a make-to-stock production system with several demand classes and lost sales. *Management Science*, 43, 1093–1103.
- Ha, AY (1997b). Stock rationing policy for a make-to-stock production system with two priority classes and backordering. *Naval Research Logistics*, 44, 457–472.
- Ha, AY (2000). Stock rationing in an $M/E_k/1$ make-to-stock queue. *Management Science*, 46, 77–87.
- Huang, B and SMR Iravani (2006). Optimal production and rationing decisions in supply chains with information sharing. *Operations Research Letters* 35, 669–676.
- Huang, B and SMR Iravani (2008). A make-to-stock system with multiple customer classes and batch ordering. *Operations Research*, 56, 1312–1320.
- Iravani, SMR, KL Luangkesorn and D Simchi-Levi (2003). On assemble-to-order systems with flexible customers. *IIE Transactions*, 35, 389–403.
- Lee, CY, W Zhou and S Wu (2011). Optimal control of a capacitated inventory system with multiple demand classes. *Naval Research Logistics*, 58(1), 43–58.
- Lin, B (2008). Dynamic Revenue Management of Airline and Production Systems. PhD Dissertation, Nanyang Technological University, Singapore.
- Lippman, S (1975). Applying a new device in the optimization of exponential queuing systems. *Operations Research*, 23, 687–710.
- Liu, S, M Song, KC Tan and C Zhang (2015). Dynamic inventory rationing among multiple classes with stochastic demands and backordering. *European Journal of Operational Research*, 244(1), 153–163.
- Nahmias, S and WS Demmy (1981). Operating characteristics of an inventory system with rationing. *Management Science*, 27, 1236–1245.
- Pang, Z, H Shen and TCE Cheng (2014). Inventory rationing in a make-to-stock system with batch production and lost sales. *Production and Operations Management*, 23(7), 1243–1257.
- Papier, F and UW Thonemann (2010). Capacity rationing in stochastic rental systems with advance demand information. *Operations Research*, 58, 274–288.
- Puterman, ML (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New Jersey: John Wiley & Sons, Inc.
- Rajagopalan, S and AC Soteriou (1994). Capacity acquisition and disposal with discrete facility sizes. *Management Science*, 40, 903–917.

- Sethi, AK and SP Sethi (1990). Flexibility in manufacturing: A survey. *International Journal of Flexible Manufacturing Systems*, 2, 512–523.
- Teunter, RH and WK Klein Haneveld (2008). Dynamic inventory rationing strategies for inventory systems with two demand classes, Poisson demand and backordering. *European Journal of Operational Research*, 190(1), 156–178.
- Tomlin, B (2006). On the value of mitigation and contingency strategies for managing supply chain disruption risks. *Management Science*, 52(5), 639–657.
- Topkis, DM (1968). Optimal ordering and rating policies in a nonstationary dynamic inventory model with n demand classes. *Management Science*, 15, 160–176.
- Turgay, Z, F Karaesmen and EL Örmeci (2015). A dynamic inventory rationing problem with uncertain demand and production rates. *Annals of Operations Research*, 231, 207–228.
- Veinott, AF (1965). Optimal policy in dynamic, single product nonstationary inventory model with several demand classes. *Operations Research*, 13, 761–778.
- Wang, D and O Tang (2014). Dynamic inventory rationing with mixed backorders and lost sales. *International Journal of Production Economics*, 149, 56–67.
- Xu, J, S Chen, B Lin and R Bhatnagar (2010). Optimal production and rationing policies of a make-to-stock production system with batch demand and backordering. *Operations Research Letters*, 38, 231–235.
- Xu, J, A Serrano and B Lin (2017). Optimal production and rationing policy of two-stage tandem production systems. *International Journal of Production Economics*, 185, 100–112.
- Zhao, H, V Deshpande and JK Ryan (2005). Inventory sharing and rationing in decentralized dealer networks. *Management Science*, 51, 531–547.

Biography

Bing Lin is an Assistant Professor in the Department of Logistics at Jiangsu Normal University. He received his PhD degree in Operations Management from Nanyang Business School at Nanyang Technological University in 2008. He has publications in *Operations Research Letters*, *European Journal of Operational Research*, *Computers & Operations Research*, *International Journal of Production Economics* and *Journal of Revenue & Pricing Management*. His research interests include modelling and analysis of stochastic systems with applications in logistics and supply chain management, production and inventory control and Revenue Management.

Shaoxiang Chen is a Professor at Nanyang Business School of Nanyang Technological University. He received his BS (Mathematics) degree from Nanjing University, China in 1982 and his MBA and PhD degrees in 1987 and 1993 from the Catholic University of Leuven, Belgium. Dr. Chen has been the Guest Professor of Executive MBA Programs at many Chinese Universities like Tsinghua University, Shanghai Jiao Tong University, and etc. He also has many years of consulting experience and is officially appointed Consultant by many Chinese municipal governments. He specializes in e-Business, Supply Chain Strategy and Design, Logistics Management, System Modelling and Optimization and has many publications in top journals like *Operations Research*, *Management Science*, *Naval Research Logistics* and *IIE Transactions*.

Yi Feng is an Associate Professor in the Department of Management Science and Electronic Commerce at the University of Science and Technology of China. He received his PhD degree in supply chain management from the Engineering School of Nanyang Technological University in 2006. He has published papers in journals including Operations Research Letters, International Journal of Production Economics and Computers & Operations Research. His research includes logistics management, supply chain management and operations management.

Jianjun Xu is an Associate Professor of the International Business College and Institute of Supply Chain Analytics at Dongbei University of Finance & Economics, China. He received his PhD degree in Operations Management from Nanyang Business School at Nanyang Technological University in 2010. His research interests include modelling and analysis in supply chain management and production/inventory control system. His work has been published in journals such as Naval Research Logistics, Operations Research Letters, European Journal of Operational Research and International Journal of Production Economics.