



Production, Manufacturing and Logistics

Prioritizing regular demand while reserving capacity for emergency demand

Xinxin Hu^a, Ying Li^{b,*}, Eunshin Byon^c, F Barry Lawrence^a^a College of Engineering, Texas A&M University, College Station, TX 77843, United States^b Mays Business School, Texas A&M University, College Station, TX 77843, United States^c College of Engineering, The University of Michigan, Ann Arbor, MI 48109, United States

ARTICLE INFO

Article history:

Received 8 January 2014

Accepted 3 June 2015

Available online 15 June 2015

Keywords:

Capacity allocation

Capacity reservation

Make-to-order

Demand prioritization

ABSTRACT

This research is motivated by the capacity allocation problem at a major provider of customized products to the oil and gas drilling industry. We formulate a finite-horizon, discrete-time, dynamic programming model in which a firm decides how to reserve capacity for emergency demand and how to prioritize two classes of regular demand. While regular demand can be backlogged, emergency demand will be lost if not fulfilled within the period of its arrival. Since backlogging cost accumulates over time, we find it optimal for the firm to adopt a dynamic prioritization policy that evaluates the priorities of different classes of regular demand every period. The optimal prioritization involves metrics that measure backlogging losses from various perspectives. We fully characterize the firm's optimal prioritization and reservation policy. Those characterizations shed light on managerial insights.

© 2015 Elsevier B.V. and Association of European Operational Research Societies (EURO) within the International Federation of Operational Research Societies (IFORS). All rights reserved.

1. Introduction

Capacity management is often a challenging task faced by many production managers. This research is motivated by the experience of a major provider of customized products to the oil and gas drilling industry. In recent years, demand for the company's products is fueled by the booming oil market. The rapid growth, however, brings challenges to capacity management. One of the most salient problems is backlogging. The company's records show that customers could wait as long as a year for their orders to be satisfied. As much as the company wants to take full advantage of the current demand increase, the executives of the company are hesitant to invest in capacity expansion for fear of market fluctuations. They are more interested in how to best utilize their existing capacity.

The majority of the company's demand is processed in a first-come-first-serve fashion. Some executives of the company question the validity of this approach because different classes of demand do not have the same gross margins. On the other hand, waiting costs are also a concern because the company hopes to maintain long-term relationships with its clients. Classifying demand based on both revenue and waiting cost is, therefore, necessary. Further

complicating the demand prioritization problem is the existence of emergency demand. Unexpected events occurring in the oil and gas field result in emergency orders of specialized equipments. In light of the long-term relationship with critical clients, the company wants to satisfy emergency demand as much as possible, even at the expense of delaying regular demand. Given the frequency of emergency demand, production managers find it necessary to reserve capacity for the uncertain emergency demand. Hence, there are two tasks in the company's capacity management problem. One is reserving capacity for unknown emergency demand. The other is splitting the leftover capacity between different classes of regular demand.

We formulate a discrete-time dynamic-programming model that reflects the two tasks faced by the company and fully characterize the optimal solution. In order to characterize the optimal policy, we develop various metrics that measure the potential losses resulting from backlogging regular demand. Comparisons among these metrics allow us to prioritize regular demand and determine the level of capacity reservation for emergency demand. We believe the optimal solution to our model provides guidance to the company's capacity management.

Although our research was motivated by challenges rising from the oil and gas drilling field, our model and solution approach can be applied to general industrial settings with customized products and capacity constraints. As pointed out by [Rahman and Seliger \(2013\)](#) and [Mourtzis, Doukas, Psarommatis, Giannoulis, and Michalos \(2014\)](#), demand for customized products has been

* Corresponding author. Tel.: +1 979 845 2392; fax: +1 979 845 5653.

E-mail addresses: hu.xinxin@hotmail.com (X. Hu), yli@mays.tamu.edu (Y. Li), ebyon@umich.edu (E. Byon), lawrence@entc.tamu.edu (F. Barry Lawrence).

increasing over the years and customization has become a widespread paradigm in manufacturing technology. Alexopoulos, Papakostas, Mourtzis, and Chryssolouris (2011) studies the performance of manufacturing systems with a focus on lifecycle, recognizing the existence of capacity constraints. deKoster, Le-Duc, and Roodbergen (2007) examines order picking systems, which feature both customized orders and capacity constraints.

The rest of the paper unfolds as follows. Section 2 reviews relevant literature. Section 3 explains the problem setting and formulates the model. We present preliminary analysis on demand types and capacity reservation, introduce the concept of dynamic loss, and discuss our solution approach in Section 4. Section 5 describes the optimal policy. Section 6 summarizes our numerical studies, which develop and compare a few heuristics to the optimal policy. Section 7 concludes the main text of the paper. Appendix A presents technical statements that are skipped in the main text for readability. Proofs of technical statements are included in a supplementary document.

2. Relevant literature

Our research is, essentially, about capacity allocation among multiple demand classes. This topic has been studied in a variety of contexts. We review, here, the relevant literature and explain how our work differs.

One stream of research considers demand that is always backlogged if unsatisfied. The same product is made and sold to multiple classes of customers. At any given time, one needs to decide whether the facility should produce or not. Once a product is made, one needs to decide whether it should be allocated to on-hand inventory, or any class of customers. An optimal solution should strike a balance between holding cost and various levels of backlogging costs. Ha (1997) started this line of research in a setting with two demand classes. de Vericourt, Karaesmen, and Dallery (2002), Huang and Iravani (2008), and Gayon, de Vericourt, and Karaesmen (2009) look at extensions or variations of Ha's (1997) work. Iravani, Liu, and Simchi-Levi (2012) look at a manufacturer who accepts all demand for a generic product with make-to-stock system and can accept or reject demand for customized products with make-to-order system. Both types of demand can be backlogged after being accepted in the setting of Iravani et al. (2012). One of the key differences between this line of research and ours is our consideration of emergency demand, which cannot be backlogged. The uncertain arrival of emergency demand during a period in our setting makes capacity reservation necessary in addition to allocating demand between two regular demand classes.

Another stream of research analyzes production and allocation decisions when a firm faces two classes of demand: contractual and transactional. The contractual demand must be met immediately while the firm chooses to accept or reject transactional demand. Carr and Duenyas (2000) formulate a continuous-time, make-to-stock model in which transactional orders are either accepted or rejected upon arrival. They show that the firm's optimal acceptance and production decisions are determined by an acceptance threshold curve and a production threshold curve. Frank, Zhang, and Duenyas (2003) adopt a discrete-time approach. Their contractual orders are deterministic and their transactional orders are stochastic. Unsatisfied transactional orders are lost. Frank et al. (2003) find that their optimal policy can be characterized by a modified and much more complicated (s, S) policy. Gupta and Wang (2007) formulate a discrete-time model and allow transactional demand to be satisfied within a fixed number of periods without penalty. When both contractual and transactional demand are make-to-order, Gupta and Wang (2007) show that the optimal acceptance decision is a threshold policy with the threshold being determined by the firm's capacity and the current period's contractual demand. The contractual demand in this

research stream differs from the regular demand in our research because contractual demand is fulfilled immediately or within a fixed number of periods without penalty. Our regular demand, in contrast, can be backlogged throughout the whole planning horizon and with accumulating penalties.

Wang, Liang, Sethi, and Yan (2014) consider two groups of customers: short lead-time ones who demand the product immediately and long lead-time ones whose orders can be fulfilled either immediately or in the next cycle. Backlogging costs accumulate for unsatisfied demand. The supplier needs to determine how to allocate on-hand inventory between the two groups of customers and how to allocate replenishment inventory among backlogged orders. The paper focuses on three priority rules and derive inventory commitment policies based on those rules.

Gao, Xu, and Ball (2012) consider pseudo orders that become actual orders after being confirmed and use a Markov chain model as a short-term forecast for pseudo orders. The manufacturer decides whether to accept or reject a confirmed order. Accepted orders must be fulfilled during a fixed time window and productions are constrained by both capacity and component inventory availability. Orders from different classes bring in different revenues.

Capacity allocation research in medical settings often deals with demand uncertainty (Gerchak, Gupta, & Henig, 1996; Patrick, Puterman, & Queyranne, 2008; Green, Savin, & Wang, 2006; Patrick & Puterman, 2007; Gupta & Wang, 2008; Min & Yih, 2010; Dobson, Hasija, & Pinker, 2011; Geng & Xie, 2012 etc.). Green (2012) discusses how operations research can improve healthcare delivery. Chen and Robinson (2010) and May, Spangler, Strum, and Vargas (2011) examine capacity allocations when scheduling appointments or surgeries, respectively. Ramirez-Nafarrate, Hafizoglu, Gel, and Fowler (2014) study ambulance diversion policies in order to minimize patients' average waiting times. Gerchak et al. (1996) is the closest to our research. They study how to share operating room capacity between elective surgeries and emergency surgeries. Before learning the amount of emergency demand arriving on any given day, the scheduler must decide how many elective surgeries are admitted for operation on the current day. Delayed elective surgeries result in postponed revenue collection and waiting costs. The elective surgeries are analogous to our regular demand but have just a single class. With a single class, the prioritization problem goes away.

Revenue management is an area that naturally deals with capacity allocation among multiple demand classes. It is often studied in contexts including airlines, hotels, and car rentals. McGill and van Ryzin (1999) and Talluri and van Ryzin (2004) provide detailed reviews. A key difference between our work and revenue management literature is the lost sales of unsatisfied demand in revenue management settings. Backlogged regular demand is more appropriate in our model because the company that motivates this research rarely turns away any demand.

A key element of our model is the existence of customized products. Given its popularity, customization has been the focus of a series of research in manufacturing settings. Mourtzis, Doukas, and Psarommatis (2013) proposes a genetic algorithm for mass customization. Mourtzis et al. (2014) examines a web-based platform that facilitates customization. Adaptability or flexibility is crucial to the manufacturers of customized products. Papakostas and Mourtzis (2007), Alexopoulos, Mourtzis, Papakostas, and Chryssolouris (2007), and Alexopoulos et al. (2011) examine the evaluations of system flexibility or adaptability. Makris and Chryssolouris (2010) incorporates demand uncertainty in manufacturing planning and applies their Bayesian model to the automotive industry. This series of research studies the design and evaluation of manufacturing systems, which often consist of multiple stages. Our model analyzes the tension between demand and supply, assuming away the details regarding manufacturing stages and layouts.

We will explain the technical differences between our model and those in the literature after formally introducing our model in the next section.

3. The model

We formulate a make-to-order firm's capacity allocation problem as a discrete-time dynamic programming model. The firm would like to maximize its discounted expected profits over T periods. We consider T a finite integer.

3.1. Regular demand and emergency demand

We consider two categories of demand in this paper. The key difference between the two categories of demand is whether demand can be backlogged. We define the category of demand that can be backlogged as regular demand. The regular demand in our model does not have to come from regular customers of a firm, and hence may not be contractual. We define the category of demand that will be lost if not satisfied within the period of arrival as emergency demand. Emergency demand can also come from long-term customers or one-time customers. Both regular and emergency demand are satisfied by the same production process. Hence, capacity allocation decisions have to be made.

Since regular demand is not necessarily driven by long-term contracts, the amount of regular demand arriving in each period need not be deterministic. Regular demand is not processed immediately upon arrival. At the end of each period, the accumulated regular demand throughout the period is added to the existing unsatisfied demand. We also recognize that demand from different customers or of different nature may have different cost structures. For this reason, we study two classes of regular demand, each with a potentially different set of cost parameters. Details regarding the definitions of the two classes of regular demand are provided in Sections 3.2 and 4.1.

Emergency demand was induced by urgent events and its arrival is, clearly, stochastic. The amount of emergency demand that arrives during a period can come from a single customer or multiple customers. While the amount of regular demand waiting to be processed is known at the beginning of a period, the amount of emergency demand that will arrive during a period is unknown when capacity allocation decisions are made for the period. The firm's capacity allocation policy needs to specify how much emergency demand can be accepted throughout the period.

We do not assume, automatically, that emergency demand has a higher priority than regular demand. Instead, our research is about finding the optimal priority rules.

Regardless of the category a unit of demand falls into, the demand faced by the firm is met by customized products. The same production process and capacity can be used to satisfy both demand categories. We assume that no changeovers are needed in the production process. As a result, demand from different sources or arriving at different times can be accumulated without affecting capacity utilization.

3.2. The sequence of events

The firm processes two classes of regular demand. Their arrival is random but stationary over the planning horizon. For $i = 1, 2$, we denote by D_i the (random) amount of class i regular demand arriving each period, f_i the corresponding probability density function, and F_i the corresponding cumulative probability function. $\mathbf{D} = (D_1, D_2)$ represents regular demand. Unsatisfied regular demand is backlogged. The firm also handles emergency demand, which will be lost if not satisfied upon its arrival. We denote by D_e the (random) amount of emergency demand arriving each period, f_e the corresponding

probability density function, F_e the corresponding cumulative probability function, and $\bar{F}_e = 1 - F_e$.

The total amount of demand that can be fulfilled by the firm is C units per period. For $t = 1, \dots, T$ and $i = 1, 2$, we denote by x_{it} the amount of unsatisfied class i demand at the beginning of period t . $\mathbf{x}_t = (x_{1t}, x_{2t})$ describes the state of the system and the state space is represented by $\Omega_t = \{(x_{1t}, x_{2t}) \mid x_{it} \geq 0 \text{ and } x_{1t} + x_{2t} > 0\}$. We exclude the state $(0, 0)$ because it results in a trivial optimal policy.

At the beginning of period t , the firm observes \mathbf{x}_t and then determines $\mathbf{y}_t = (y_{1t}, y_{2t})$ (how much of each class of regular demand to fulfill within period t). Obviously, y_{it} cannot exceed x_{it} for $i = 1, 2$. Furthermore, the inequality $C \geq y_{1t} + y_{2t}$ must hold. We normalize the cost of fulfilling a unit of demand using the firm's own capacity to zero. For $i = 1, 2$, each unit of fulfilled class i demand earns the firm a revenue of r_i and each unit of unfulfilled class i demand results in a backlogging cost of b_i for the period. With the decision of \mathbf{y}_t , one can infer that the firm has made a capacity reservation for emergency demand of $C - y_{1t} - y_{2t}$ units in period t .

After determining its capacity allocation for period t , the firm learns the realized value of D_e , namely the amount of emergency demand arriving in period t . The emergency demand that is fulfilled by the firm within period t is $D_e \wedge (C - y_{1t} - y_{2t})$. For each unit of emergency demand beyond $C - y_{1t} - y_{2t}$, the firm loses c_e , which is the revenue the firm earns by satisfying a unit of emergency demand.

The firm learns the realized values of D_1 and D_2 at the end of period t . The state of the system at the beginning of the next period becomes $\mathbf{x}_{t+1} = (x_{1t} - y_{1t} + D_1, x_{2t} - y_{2t} + D_2)$. Note that we label the starting period as period T and the ending period as period 1.

3.3. The MDP formulation

Upon observing its system state \mathbf{x}_t , the firm makes a capacity reservation of $C - y_{1t} - y_{2t}$ units for emergency demand and fulfills y_{it} units of regular demand from class i in period t , for $i = 1, 2$. These decisions can be captured by $z_{it} = x_{it} - y_{it}$, $i = 1, 2$. We refer to z_{it} as the amount of class i demand to be backlogged in period t . In the MDP formulation below, we let the vector $\mathbf{z}_t = (z_{1t}, z_{2t})$ be the firm's decision variables in period t for the convenience of analysis. The action space becomes $\mathcal{Z}_t(\mathbf{x}_t) = \{(z_{1t}, z_{2t}) \mid x_{it} \geq z_{it} \geq 0, z_{1t} + z_{2t} \geq x_{1t} + x_{2t} - C\}$. Note that $C - \sum_{i=1}^2 (x_{it} - z_{it})$ is the level of capacity reservation for emergency demand because $x_{it} - z_{it}$ is the amount of class i regular demand to be fulfilled in period t .

The firm's expected revenue from the emergency demand can be expressed as the difference between $c_e \mathbf{E}[D_e]$ and the expected lost revenue $c_e \mathbf{E}[D_e - (C - \sum_{i=1}^2 (x_{it} - z_{it}))]^+$. Note that $c_e \mathbf{E}[D_e]$ is a constant that is not affected by either the firm's decision \mathbf{z}_t or the firm's state \mathbf{x}_t . For this reason, we omit the term $c_e \mathbf{E}[D_e]$ from the firm's profit function throughout the paper.

We denote by $G_t(\mathbf{x}_t)$ the firm's maximal, expected profit function from period t to period 1 (the end of the planning horizon). Using α as the discount factor, the firm's maximal expected profit can be expressed as follows:

$$G_t(\mathbf{x}_t) = \max_{\mathbf{z}_t \in \mathcal{Z}_t} \left\{ \sum_{i=1}^2 (r_i (x_{it} - z_{it}) - b_i z_{it}) - c_e \mathbf{E} \left[D_e - \left(C - \sum_{i=1}^2 (x_{it} - z_{it}) \right) \right]^+ + \alpha \mathbf{E} G_{t-1}(\mathbf{z}_t + \mathbf{D}) \right\}$$

with the boundary condition: $G_0(\mathbf{x}_0) \equiv 0$.

In order to facilitate our analysis, we introduce the following notations that are related to state variables and profit functions:

- $X_t = x_{1t} + x_{2t}$: the aggregate amount of unsatisfied regular demand at the beginning of period t ;
- $Z_t = z_{1t} + z_{2t}$: the total amount of regular demand to be backlogged in period t ;

- $V_t(\mathbf{z}_t, X_t) = \sum_{i=1}^2 -(r_i + b_i)z_{it} - c_e \mathbf{E}[D_e - (C - X_t + Z_t)]^+ + \alpha \mathbf{E}G_{t-1}(\mathbf{z}_t + \mathbf{D})$;
- $\bar{V}_t(\mathbf{z}_t, \mathbf{x}_t) = \sum_{i=1}^2 r_i x_{it} + V_t(\mathbf{z}_t, X_t)$.

With the above newly defined notations, the maximal expected profit function $G_t(\mathbf{x}_t)$ can be rewritten as:

$$G_t(\mathbf{x}_t) = \max_{\mathbf{z}_t \in Z_t} \bar{V}_t(\mathbf{z}_t, \mathbf{x}_t) = \max_{\mathbf{z}_t \in Z_t} \left\{ \sum_{i=1}^2 r_i x_{it} + V_t(\mathbf{z}_t, X_t) \right\}.$$

The new expression reveals our motivation to introduce $V_t(\mathbf{z}_t, X_t)$, which depends on X_t , the aggregate amount of unsatisfied regular demand. Since $\sum_{i=1}^2 r_i x_{it}$ is not affected by the decision vector \mathbf{z}_t , maximizing $\bar{V}_t(\mathbf{z}_t, \mathbf{x}_t)$ is equivalent to maximizing $V_t(\mathbf{z}_t, X_t)$. Such recognition implies that X_t will play an important role as we search for the optimal solution. As a matter of fact, we will establish a number of thresholds on X_t and apply them when developing the optimal policy in Section 5.

Before presenting our analysis and solution, we would like to briefly explain, here, how our research differs from those in the literature from a modeling perspective:

- Capacity reservations are made for emergency demand before its arrival. We study this setting because a major challenge faced by the managers in the real world is the uncertainty associated with emergency demand. How to reserve capacity for emergency demand is a key question we would like to address in this research.
- The capacity allocation problem we analyze involves sharing capacity between two classes of regular demand in addition to making reservations for emergency demand. In existing literature, when there are multiple classes of regular demand, the class of regular demand with the highest backlogging or lost-sales cost is naturally assigned the highest priority. We consider both revenue and backlogging costs, which can have the same or different ordinal ranks among various classes of regular demand. While revenue is collected only at the time of order fulfillment, backlogging costs accumulate over multiple periods until the demand is satisfied. The class of regular demand with a lower revenue may have a higher accumulated backlogging cost if unfulfilled for multiple periods. Consequently, the priority level of any given class of regular demand need not remain the same over the entire planning horizon. We develop a number of definitions that measure backlogging losses, and subsequently, the optimal capacity allocation is based on those measurements.
- Regular demand is never turned away and unsatisfied regular demand is backlogged. From the analytical perspective, a backlog setting is more difficult to analyze than a lost-sales setting because of the need to keep track of unsatisfied demand. Emergency demand cannot be backlogged in our model. It will be lost if not satisfied within the period of its arrival.

4. Preliminary analysis and solution approach

The existence of multiple demand classes invites prioritization rules. Appropriate classification of demand is key to the development of prioritization rules. Our preliminary analysis starts with a close look at demand types (Section 4.1). Building upon notations introduced in Section 4.1, we develop more metrics on backlogging losses and capacity reservation in Sections 4.2 and 4.3. We present in Section 4.4 the solution approach, which utilizes notations built upon demand types.

4.1. The period and limit types

4.1.1. Definitions

The firm’s decision should be driven by how revenues are earned and how costs are incurred. Satisfying a unit of class i demand earns the firm a revenue of r_i while backlogging a unit of class i demand costs b_i per period. When prioritizing the two classes of regular demand, r_i ’s and b_i ’s should play an important role.

It would not be surprising if a firm values demand with higher-margins more than demand with lower-margins. In this case, the class of regular demand with a higher r_i is also the class of regular demand with a higher b_i . There is, however, another possibility. A firm may have long-term customers whose business is essential to the firm’s reputation. In order to maintain the relationship with those long-term customers, the backlogging cost associated with their demand could be higher even if the demand generates a lower unit revenue. Taking into consideration both possibilities, we do not assume that the class of regular demand with a higher r_i is automatically the class with a higher b_i . Note that b_i reflects, essentially, the firm’s relationship with a customer. A higher b_i implies a closer and more valuable relationship, which can be driven by profit margins, or other concerns like long-term business opportunities, market penetrations, reputations, etc.

Because of the possibilities discussed above, neither r_i nor b_i alone is likely to be a good indicator of the priority that should be assigned to class i of regular demand. Another complicating factor is that the backlogging cost accumulates if the corresponding regular demand remains unfulfilled through multiple periods. We need to, therefore, look for better ways to classify regular demand.

At period t , if the firm chooses not to fulfill a unit of class i regular demand throughout the rest of the planning horizon, the total loss the firm incurs is the sum of a loss of revenue (r_i) and an accumulation of discounted backlogging costs ($b_i \sum_{k=0}^{t-1} \alpha^k$) from period t to the end of the planning horizon. This sum also represents the maximal loss that the firm can possibly suffer over the rest of the planning horizon from its decision of not fulfilling a unit of class i demand in period t . We believe this sum is a meaningful criterion for the prioritization of regular demand and define the following notations to represent it:

- $n_{it} = b_i \sum_{k=0}^{t-1} \alpha^k$ and $m_{it} = r_i + n_{it}$.

Note that m_{it} is the aforementioned sum. Since it sets an upper bound on the amount of loss the firm can possibly incur thereafter due to the decision of backlogging a unit of class i demand in period t , we refer to m_{it} as class i ’s Maximal Loss (ML) in period t . The metric m_{it} is period-dependent because it includes a period-dependent part: n_{it} , which is the accumulated backlogging costs throughout the rest of the planning horizon. The closer period t is to the end of the planning horizon, the smaller n_{it} and m_{it} are.

Considering that $r_1 > r_2$ can co-exist with $b_1 < b_2$, the firm may have $m_{1t} \geq m_{2t}$ in one period but $m_{1(t-1)} < m_{2(t-1)}$ in the next period. For this reason, the priorities indicated by m_{it} ’s are only applicable to period t . The dynamic nature of m_{it} prompts us to consider a stationary indicator: m_i , which is the limit of m_{it} as t goes to infinity as follows:

- $n_i = \lim_{t \rightarrow \infty} n_{it}$ and $m_i = \lim_{t \rightarrow \infty} m_{it} = r_i + n_i$.

With these notations, we formally introduce two ways of classifying regular demand: period type and limit type, in Definition 4.1.

Definition 4.1. For $i = 1, 2$ and $j = 3 - i$, class i of regular demand has a period type of *High* in period t if $m_{it} \geq m_{jt}$, and a period type of *Low* otherwise. The class of demand with a period type of High (or Low) is referred to as the PH (or PL) demand. Class i of regular demand has a limit type of *High* if $m_i \geq m_j$, and a limit type of *Low* otherwise. The class of demand with a limit type of High (or Low) is referred to as the LH (or LL) demand.

The difference between period type and limit type is essentially whether the impact of t , the current period, is considered. Naturally, the definition that takes into consideration the impact of t is about period types, and the definition that brings limit into consideration is about limit types.

Note that, whether it is period type or limit type, the class of regular demand with a higher criterion value is referred to as the high type.

4.1.2. Properties

Next we investigate the properties of m_{it} and m_i , which are at the core of our demand classifications.

An increase in either unit revenue (r_i) or unit backlogging cost (b_i) results in both a higher m_{it} and a higher m_i . While r_i has the same impact on both m_{it} and m_i , m_{it} and m_i increase at different rates as b_i goes up. This is due to the fact that backlogging costs accumulate over periods in which the corresponding demand remains unsatisfied. Therefore, whether t is a part of the criterion value matters. It might be natural for one to expect the class of regular demand with a higher limit type to have a higher period type as well. However, given how b_i impacts m_{it} and m_i , it is also possible for the class of regular demand with a higher limit type to have a lower period type. **Lemma 4.2** below quantifies when period type and limit type are consistent and when they are not.

Lemma 4.2. For $i = 1, 2$ and $j = 3 - i$, there exists $T^0 \geq 0$ such that given $m_i \geq m_j$, (i) if $T \leq T^0$, then $m_{it} < m_{jt}$ for $t = 1, 2, \dots, T$; (ii) if $T > T^0$, then $t \leq T^0 \Rightarrow m_{it} < m_{jt}$ and $T^0 < t \leq T \Rightarrow m_{it} \geq m_{jt}$.

Lemma 4.2 establishes a period threshold T^0 . Basically, if there are enough number of periods left before the end of the planning horizon (i.e. $t > T^0$), then the regular demands' period and limit types are consistent; otherwise they are inconsistent.

As noted earlier, b_i represents the firm's relationship with a customer. Consider a customer whose demand carries a relatively smaller unit revenue r_i . If this customer is not as valued to the firm as the source of the other demand, then this customer's demand has a relatively smaller b_i and is always of the low type both period-wise and limit-wise. If this customer has a relatively bigger b_i , then it is possible for the corresponding demand to be of the high type both period-wise and limit-wise. First of all, b_i has to be big enough for m_i (the limit of m_{it}) to exceed m_j . This way, the customer's demand is of the limit high (LH) type. **Lemma 4.2** shows that an LH demand ($m_i \geq m_j$), however, is not guaranteed to also be a period high (PH) demand. In a period that is faraway enough from the end of the planning horizon, b_i can accumulate over enough number of periods to compensate for the relatively lower r_i and result in a relatively bigger m_{it} . Depending on the relative sizes of r_i , b_i , r_j and b_j , the threshold period T^0 can be any integer between 0 and infinity. The bigger T^0 is, the more periods it takes for b_i to accumulate to a point that the class of demand with a bigger b_i is period high (PH). In other words, given a planning horizon that is long enough, if a customer is important/valuable enough to the firm (i.e. has LH demand), there are definitely periods in which this customer's demand is of the period high (PH) type.

Based on the definitions of period and limit types, we are able to identify, in later sections, conditions under which the PH demand dominates the PL demand. Intuitively, the dominant demand type should be assigned a higher priority. The conditions for the existence of a dominant demand, thus, help us describe the optimal policies.

4.2. Dynamic losses

What m_{it} represents is the maximal loss from backlogging a unit of class i demand in period t . The actual backlogging loss of a unit of class i demand depends on the amount of unsatisfied demand from

each regular demand class, and the firm's capacity allocation decisions in future periods. Specifically, when z_{it} units of class i demand and z_{jt} units of class j demand are currently being backlogged, postponing the fulfillment of an additional unit of class i demand results in a loss of revenue r_i and a backlog penalty of b_i in period t . Note that $r_i + b_i = m_{i1}$. The marginal impact of the action on profits from future periods can be expressed by $\alpha \mathbf{E}(G_{t-1})'_i(\mathbf{z}_t + \mathbf{D})$, where α is the discount factor and $\mathbf{E}(G_{t-1})'_i(\mathbf{z}_t + \mathbf{D}) = \frac{\partial \mathbf{E}G_{t-1}(\mathbf{z}_t + \mathbf{D})}{\partial z_{it}}$. With this understanding, we define

$$\bar{m}_{it}(z_{it}, z_{jt}) := m_{i1} - \alpha \mathbf{E}(G_{t-1})'_i(\mathbf{z}_t + \mathbf{D})$$

and refer to $\bar{m}_{it}(z_{it}, z_{jt})$ as the Dynamic Loss (DL) for class i in period t .

We further define the Starting Dynamic Loss (SDL) \bar{m}_{it}^S and the Limiting Dynamic Loss (LDL) \bar{m}_{it}^L . By setting (z_{it}, z_{jt}) to specific values, we have:

$$\bar{m}_{it}^S = \bar{m}_{it}(0, 0),$$

$$\bar{m}_{it}^L = \lim_{z_{jt} \rightarrow \infty} \bar{m}_{it}(0, z_{jt}).$$

Since the best the firm can do with regard to regular demand alone is not to backlog at all, we know that \bar{m}_{it}^S does not exceed \bar{m}_{it}^L . A more thorough study of dynamic losses is in **Lemma 4.3**, which examines how \bar{m}_{it} changes as decision variables change and how one class's dynamic losses compare to the other class's.

Lemma 4.3. For $i = 1, 2$ and $j = 3 - i$, $\bar{m}_{it}(z_{it}, z_{jt})$ has the following properties: 1. $\frac{\partial \bar{m}_{it}}{\partial z_{it}} \geq \frac{\partial \bar{m}_{it}}{\partial z_{jt}} \geq 0$ and $\frac{\partial \bar{m}_{it}}{\partial z_{it}} \geq \frac{\partial \bar{m}_{jt}}{\partial z_{it}}$; 2. for any $z \geq 0$, $\bar{m}_{it}(z_{it}, z_{jt} + z) \leq \bar{m}_{it}(z_{it} + z, z_{jt})$; 3. $\bar{m}_{it}^S \leq \bar{m}_{it}^L \leq m_{it}$; 4. Given $m_{it} \geq m_{jt}$, if $m_i \geq m_j$ then $\bar{m}_{it}^S \geq \bar{m}_{jt}^S$, otherwise, $\bar{m}_{it}^L \geq m_{jt} \Rightarrow \bar{m}_{it}^S \geq \bar{m}_{jt}^S$.

4.3. Minimal capacity reservations

The metrics m_{it} 's also lead to bounds on capacity reservation, which depends crucially on how the unit lost revenue c_e compares to the loss from backlogging a unit of regular demand. Although the backlogging loss is affected by when the backlogged demand is eventually fulfilled, it is bounded from above by m_{it} , which is the maximal loss from period t and onward. If the firm had only one class of regular demand and needed to allocate its capacity between emergency demand and the regular demand, the Newsvendor logic could be applied in capacity reservation. Adopting m_{it} as the overage cost and $(c_e - m_{it})^+$ the underage cost, we define \bar{C}_{it} as follows. For $i = 1, 2$,

$$\bar{C}_{it} = \begin{cases} F_e^{-1}\left(1 - \frac{m_{it}}{c_e}\right), & \text{if } c_e \geq m_{it}; \\ 0, & \text{if } c_e < m_{it}. \end{cases}$$

Since m_{it} is a unit of class i demand's ML (maximal loss) thereafter if backlogged in period t , \bar{C}_{it} is the minimal capacity reservation for emergency demand from class i 's perspective. Note that when $c_e < m_{it}$, \bar{C}_{it} is forced to zero. In that case, there should be no reservation for emergency demand at all if class i constitutes the only class of regular demand.

Since the value of m_{it} depends on the period t in question, the length of the planning horizon also affects how c_e compares to m_{it} . We show in **Lemma 4.4** the existence of horizon length thresholds that allow m_{it} to overtake c_e .

Lemma 4.4. For $i = 1, 2$, there exist a T_i^0 that satisfies the following conditions.

1. If $T \leq T_i^0$, then $c_e \geq m_{it}$ for all $t = 1, \dots, T$;
2. If $T > T_i^0$, then $t \leq T_i^0 \Rightarrow c_e \geq m_{it}$ and $T_i^0 < t \leq T \Rightarrow c_e < m_{it}$.

4.4. Solution approach

With notations and preliminary analysis introduced in earlier sections, we are ready to present our solution approach.

The goal of the firm is to maximize the function $V_t(\mathbf{z}_t, X_t)$ with respect to a two-dimension vector \mathbf{z}_t , subject to the constraint $\mathbf{z}_t \in \mathcal{Z}_t(\mathbf{x}_t)$. Recall that $V_t(\mathbf{z}_t, X_t) = \sum_{i=1}^2 -(r_i + b_i)z_{it} - c_e \mathbf{E}[D_e - (C - X_t + Z_t)]^+ + \alpha \mathbf{E}G_{t-1}(\mathbf{z}_t + \mathbf{D})$ and $\mathcal{Z}_t(\mathbf{x}_t) = \{z_{1t}, z_{2t} \mid 0 \leq z_{it} \leq x_{it}, X_t - C \leq Z_t\}$.

Our analysis starts with a relaxation of the constraints on the decision variables. Specifically, we relax $\mathcal{Z}_t(\mathbf{x}_t)$ to

$$\mathcal{Z}_t^r(X_t) = \{z_{1t}, z_{2t} \mid 0 \leq z_{it} \leq X_t, X_t - C \leq Z_t \leq X_t\},$$

With the relaxation, the decision variables z_{1t} and z_{2t} are both constrained by the aggregate amount of unsatisfied demand X_t , rather than by x_{1t} and x_{2t} , respectively. The relaxation reduces the dimensions of the constraints and therefore simplifies the analysis.

Under the relaxation, we would like to establish for each regular demand class a best response function $\bar{z}_{it}(z_{jt}, X_t)$, corresponding to a given X_t . Naturally, we need to make use of the first order derivatives of $V_t(\mathbf{z}_t, X_t)$. For $i = 1, 2$ and $j = 3 - i$, we have

$$v_{it}(z_{it}, z_{jt}, X_t) = \frac{\partial V_t}{\partial z_{it}}(\mathbf{z}_t, X_t) = -m_{i1} + c_e \bar{F}_e(Z_t - X_t + C) + \alpha \mathbf{E}(G_{t-1})'_i(\mathbf{z}_t + \mathbf{D}),$$

where $m_{i1} = r_i + b_i$ and $\mathbf{E}(G_{t-1})'_i(\mathbf{z}_t + \mathbf{D}) := \frac{\partial \mathbf{E}G_{t-1}(\mathbf{z}_t + \mathbf{D})}{\partial z_{it}}$.

Now we formally define $\bar{z}_{it}(z_{jt}, X_t)$ as follows:

$$\bar{z}_{it}(z_{jt}, X_t) = \underset{[X_t - (C - \bar{c}_{it}) - z_{jt}]^+ \leq z_{it} \leq X_t - z_{jt}}{\text{arg min}} |v_{it}(z_{it}, z_{jt}, X_t)|. \quad (1)$$

Note that z_{it} : the amount of class i demand to be backlogged in period t , cannot exceed $X_t - z_{jt}$ because X_t is the aggregate amount of unsatisfied regular demand in period t . On the other hand, if the capacity reservation is \bar{c}_{it} : the minimal level from class i 's perspective, $C - \bar{c}_{it}$ units of capacity will be used on fulfilling regular demand. That is, at least $[X_t - (C - \bar{c}_{it})]^+$ units of regular demand should be backlogged in period t . Given that z_{jt} units of class j demand are to be backlogged in period t , z_{it} must be bounded from below by either $X_t - (C - \bar{c}_{it}) - z_{jt}$ or 0, whichever is bigger.

The intersection of $\bar{z}_{it}(z_{jt}, X_t)$ and $\bar{z}_{jt}(z_{it}, X_t)$ gives us an equilibrium optimal:

$$\mathbf{z}_t^e(X_t) = \underset{\mathbf{z}_t \in \mathcal{Z}_t^e(X_t)}{\text{arg min}} |\bar{v}_{it}(z_{it}, z_{jt})|, \quad (2)$$

where

$$\mathcal{Z}_t^e(X_t) = \left\{ (z_{it}, z_{jt}) \mid \begin{cases} \bar{z}_{it}(z_{jt}, X_t) - z_{it} = 0 \\ \bar{z}_{jt}(z_{it}, X_t) - z_{jt} = 0 \end{cases} \right\}$$

and

$$\begin{aligned} \bar{v}_{it}(z_{it}, z_{jt}) &= v_{it}(z_{it}, z_{jt}, X_t) - v_{jt}(z_{it}, z_{jt}, X_t) \\ &= -\{m_{i1} - \alpha \mathbf{E}(G_{t-1})'_i(\mathbf{z}_t + \mathbf{D})\} \\ &\quad + \{m_{j1} - \alpha \mathbf{E}(G_{t-1})'_j(\mathbf{z}_t + \mathbf{D})\}. \end{aligned}$$

Recall that we obtain the equilibrium optimal \mathbf{z}_t^e by relaxing the action space from $\mathcal{Z}_t(\mathbf{x}_t)$ to $\mathcal{Z}_t^r(X_t)$. By imposing the constraint $\mathbf{z}_t \leq \mathbf{x}_t$ on \mathbf{z}_t^e , we can return to the original action space and the original optimization problem.

Denote by \mathbf{z}_t^* the optimal solution. If $\mathbf{z}_t^e(X_t) \leq \mathbf{x}_t$, then $\mathbf{z}_t^*(\mathbf{x}_t) = \mathbf{z}_t^e(X_t)$. Otherwise, at least one of the dimensions in \mathbf{z}_t^e needs to be truncated to its counterpart in \mathbf{x}_t . For example, if $z_{it}^e(X_t) > x_{it}$, then the optimal solution must have $z_{it}^* = x_{it}$. In this case, not processing any unit of class i demand in period t is the optimal decision.

Ultimately, we have

$$\begin{aligned} \mathbf{z}_{it}^*(\mathbf{x}_t) &= \mathbf{1}(\mathbf{x}_t \geq \mathbf{z}_t^e) \mathbf{z}_{it}^e(X_t) + \mathbf{1}(x_{it} < z_{it}^e) x_{it} + \mathbf{1}(x_{jt} < z_{jt}^e) \bar{z}_{it}(x_{jt}, X_t), \\ \mathbf{z}_{jt}^*(\mathbf{x}_t) &= \mathbf{1}(\mathbf{x}_t \geq \mathbf{z}_t^e) \mathbf{z}_{jt}^e(X_t) + \mathbf{1}(x_{it} < z_{it}^e) \bar{z}_{jt}(x_{it}, X_t) + \mathbf{1}(x_{jt} < z_{jt}^e) x_{jt}. \end{aligned} \quad (3)$$

The technical statements that lead to the validation of $\bar{z}_{it}(z_{jt}, X_t)$, $\mathbf{z}_t^e(X_t)$, and $\mathbf{z}_t^*(\mathbf{x}_t)$ are included in [Appendix A.3](#). The proofs of those statements involve inductions properties (listed in [Appendix A.2](#)), which should hold for any period t .

Now we have presented our solution approach. It is still unclear, though, what the optimal policy looks like. We characterize the optimal policy in [Section 5](#) through structural results that identify switching curves. Those switching curves segment the state space and speed up the search for the optimal solution.

5. The optimal policy

The optimal policy should tell the firm what to do at any possible point within the state space $\Omega_t = \{(x_{1t}, x_{2t}) \mid x_{it} \geq 0 \text{ and } x_{1t} + x_{2t} > 0\}$. Our first step toward describing the optimal policy is defining four non-overlapping regions within the state space Ω_t , which is the first quadrant excluding the origin. It is not surprising that for states that are next to each other, optimal actions are of the same or similar characteristics, suggesting that the optimal actions taken by states belonging to the same region share a common structure. By describing the characteristics of the optimal actions, we can distinguish one region from another.

Below is our definition of the four regions.

Definition 5.1. We can define four regions in Ω_t , each featuring a unique pattern for $\mathbf{z}_t^*(\mathbf{x}_t)$:

1. Full-Reservation region Ω_t^F : when $\mathbf{x}_t \in \Omega_t^F$, all C units of capacity are reserved for emergency demand. That is, $\mathbf{z}_t^*(\mathbf{x}_t) = \mathbf{x}_t$.
2. Only-High region Ω_t^H : when $\mathbf{x}_t \in \Omega_t^H$, no capacity is allocated to the PL demand, and the PH demand always gets some capacity. Mathematically, if $m_{it} \geq m_{jt}$, we have $z_{it}^* = \bar{z}_{it}(x_{jt}, X_t) < x_{it}$ if $x_{it} > 0$ and $z_{jt}^* = x_{jt}$.
3. Only-Low region Ω_t^L : when $\mathbf{x}_t \in \Omega_t^L$, no capacity is allocated to the PH demand, and the PL demand always gets some capacity. Mathematically, if $m_{it} \geq m_{jt}$, we have $z_{it}^* = \bar{z}_{it}(x_{it}, X_t) < x_{it}$ if $x_{jt} > 0$ and $z_{it}^* = x_{it}$.
4. Diagonal-Stable region Ω_t^S : when $\mathbf{x}_t \in \Omega_t^S$, both classes of regular demand are assigned some capacity, and for any given X_t , the total amount of regular demand to be backlogged in period t is a constant. Mathematically, $\mathbf{z}_t^*(\mathbf{x}_t) = \mathbf{z}_t^e(X_t)$.

Next we explain how to determine which region any given state (x_{1t}, x_{2t}) should be in. [Section 5.1](#) offers a general outlook of the landscape. [Sections 5.2](#) and [5.3](#) provide refined descriptions of the regions under different scenarios. [Section 5.4](#) summarizes the optimal policy.

5.1. General outlook

Recall that \bar{c}_{1t} and \bar{c}_{2t} represent the minimal capacity that must be reserved for emergency demand from the perspective of class 1 and class 2 demand, respectively. We compare the firm's capacity C to \bar{c}_{1t} and \bar{c}_{2t} in this section. Different ordinal ranks lead to different optimal solutions. We present the general results in [Theorems 5.2, 5.3, and 5.5](#).

5.1.1. When the firm's capacity is below at least one of the minimal capacity reservation levels

If the firm's capacity is so tight that it covers neither \bar{c}_{1t} nor \bar{c}_{2t} , the firm has to reserve all its capacity for emergency demand. [Theorem 5.2](#) below formalizes what it means for the firm to be tight on capacity.

Theorem 5.2. If $C \leq (\bar{C}_{1t} \wedge \bar{C}_{2t})$, then $\Omega_t = \Omega_t^F$; otherwise $\Omega_t^F = \emptyset$.

The Full-Reservation region calls for the devotion of the firm's full capacity to emergency demand. Theorem 5.2 shows that when the condition $C \leq (\bar{C}_{1t} \wedge \bar{C}_{2t})$ is met, the Full-Reservation region is the same as the state space, and therefore the firm's optimal decision is no longer affected by the state vector $\mathbf{x}_t = (x_{1t}, x_{2t})$. Theorem 5.2 also indicates that the Full-Reservation region disappears completely once the firm's capacity exceeds the minimum of \bar{C}_{1t} and \bar{C}_{2t} . That is, as the firm's capacity gets larger, some regular demand will be fulfilled.

Next we examine in Theorem 5.3 what happens when the firm's capacity is between $\bar{C}_{1t} \wedge \bar{C}_{2t}$ and $\bar{C}_{1t} \vee \bar{C}_{2t}$. We find that no PL demand will be satisfied in this case. Recall that the PL demand has a smaller ML (Maximal Loss). The definition of minimal capacity reservation implies that the class of regular demand with a smaller ML has a larger minimal reservation. That is, the minimal capacity reservation from the PL demand's perspective is $\bar{C}_{1t} \vee \bar{C}_{2t}$. When the firm's capacity C is smaller than $\bar{C}_{1t} \vee \bar{C}_{2t}$, letting the PL demand wait costs the firm less than losing emergency demand.

We show in Theorem 5.3 that when the firm's capacity is between $\bar{C}_{1t} \wedge \bar{C}_{2t}$ and $\bar{C}_{1t} \vee \bar{C}_{2t}$, the firm will first reserve $\bar{C}_{1t} \wedge \bar{C}_{2t}$ units of capacity for emergency demand, then allocate the remaining capacity between the PH demand and emergency demand. If class i is the PH demand, the allocation depends on two thresholds: X_{it}^P (as the low threshold) and X_{it}^C (as the high threshold). Their mathematical definitions are in Appendix A.3. If the amount of unsatisfied PH demand is smaller than X_{it}^P , the remaining capacity is enough to satisfy all the PH demand and provide additional reservation for emergency demand. If the amount of unsatisfied PH demand is larger than X_{it}^C , all the remaining capacity will be used on the PH demand and there will still be PH demand remaining unfulfilled. If the amount of unsatisfied PH demand is between X_{it}^P and X_{it}^C , out of $C - \bar{C}_{1t} \wedge \bar{C}_{2t}$ units of capacity, some will be reserved for emergency demand and the rest will be used on the PH demand, which will again not be fulfilled completely.

Theorem 5.3. Suppose $(\bar{C}_{1t} \wedge \bar{C}_{2t}) < C \leq (\bar{C}_{1t} \vee \bar{C}_{2t})$. We have $\Omega_t = \Omega_t^H$. Furthermore, if class i of regular demand is the PH demand and class j is the PL demand in period t , that is, $m_{it} \geq m_{jt}$, the following statements will hold. (1) If $0 \leq x_{it} \leq X_{it}^P$, \bar{C}_{it} or more units of capacity are reserved for emergency demand, and all the PH demand are satisfied; (2) if $X_{it}^P \leq x_{it} \leq X_{it}^C$, \bar{C}_{it} or more units of capacity are reserved for emergency demand, and only $x_{it} - \bar{z}_{it}(x_{jt}, X_t)$ units of the PH demand are satisfied, with $\bar{z}_{it}(x_{jt}, X_t) \geq x_{it} - (C - \bar{C}_{it})$, $\frac{\partial \bar{z}_{it}}{\partial x_{it}} \in [0, 1]$, and $\frac{\partial \bar{z}_{it}}{\partial x_{jt}} = 0$; (3) if $X_{it}^C \leq x_{it}$, exactly \bar{C}_{it} units of capacity are reserved for emergency demand while the rest of the capacity is used on the PH demand but can only satisfy a fraction of it.

5.1.2. Thresholds On X_t

When the firm's capacity exceeds $\bar{C}_{1t} \vee \bar{C}_{2t}$, the minimal capacity reservation is met from the perspectives of both classes of regular demand. It is not surprising that we can always find states at which both the PH and PL demand are at least partially fulfilled. That is, the Diagonal-Stable region will not be an empty set when the firm's capacity exceeds $\bar{C}_{1t} \vee \bar{C}_{2t}$. Recall that the optimal solution in the Diagonal-Stable region is $\mathbf{z}_t^*(\mathbf{x}_t) = \mathbf{z}_t^e(X_t)$ (as stated in the definition of the Diagonal-Stable region), which depends on X_t . For this reason, we establish four thresholds on X_t .

- X_t^{FS} is the aggregate amount of unsatisfied regular demand below which both classes of regular demand are Fully Satisfied. That is, $X_t \leq X_t^{FS} \Leftrightarrow \mathbf{z}_t^e(X_t) = (0, 0)$.
- X_t^{PS} is the aggregate amount of unsatisfied regular demand beyond which both classes of regular demand are Partially Satisfied. That is, $X_t > X_t^{PS} \Leftrightarrow \mathbf{z}_t^e(X_t) > (0, 0)$.
- X_t^{MR} is the aggregate amount of unsatisfied regular demand beyond which the firm makes a Minimal Reservation ($\bar{C}_{1t} \vee \bar{C}_{2t}$) for

emergency demand. That is, $X_t > X_t^{MR} \Leftrightarrow X_t - (z_{1t}^e + z_{2t}^e) = [C - (\bar{C}_{1t} \vee \bar{C}_{2t})]^+$.

- X_t^{LL} is the aggregate amount of unsatisfied regular demand beyond which the desired amount of PL demand to be backlogged is Linearly Increasing in X_t at the rate of 1. That is, when $m_{it} \geq m_{jt}$, we have $X_t \geq X_t^{LL} \Leftrightarrow \frac{\partial z_{jt}^e}{\partial X_t} = 1$ and $\frac{\partial z_{it}^e}{\partial X_t} = 0$.

In Lemma 5.4 below, we verify the existence of the above four thresholds, and establish the relationships among them.

Lemma 5.4. Given $C > (\bar{C}_{1t} \vee \bar{C}_{2t})$, we have $\frac{\partial z_{1t}^e}{\partial X_t} \in [0, 1]$, $\frac{\partial z_{2t}^e}{\partial X_t} \in [0, 1]$ and $\frac{\partial z_{1t}^e}{\partial X_t} + \frac{\partial z_{2t}^e}{\partial X_t} < 1$. The four thresholds $\{X_t^{FS}, X_t^{PS}, X_t^{MR}, X_t^{LL}\}$ can be identified, and have the following properties when $\bar{m}_{it}^S \geq \bar{m}_{jt}^S$:

1. If $m_{it}^L \geq m_{jt}$, then $0 < X_t^{FS} \leq X_t^{MR} = X_t^{LL} < X_t^{PS} = \infty$, $z_{it}^e(X_t) = 0$, and $\bar{v}_{it}(z_{it}^e, z_{jt}^e) \leq 0$.
2. If $\bar{m}_{it}^L < m_{jt}$ and $c_e > (\bar{m}_{it} \wedge \bar{m}_{jt})$, then $0 < X_t^{FS} \leq X_t^{PS} \leq X_t^{MR} = X_t^{LL} < \infty$.
3. If $\bar{m}_{it}^L < m_{jt}$ and $c_e \leq (\bar{m}_{it} \wedge \bar{m}_{jt})$, then there exists $\bar{m}_{it}^E \in [\bar{m}_{it}^S, \bar{m}_{it}^L]$ such that if $c_e \leq \bar{m}_{it}^E$ then $0 < X_t^{FS} \leq X_t^{MR} \leq X_t^{PS} \leq X_t^{LL} < \infty$; otherwise, $0 < X_t^{FS} < X_t^{PS} \leq X_t^{MR} \leq X_t^{LL} < \infty$.
4. Given $\bar{m}_{it}^L < m_{jt}$, if $X_t < X_t^{PS}$, then $z_{it}^e(X_t) = 0$ and $\bar{v}_{it}(z_{it}^e, z_{jt}^e) \leq 0$; otherwise $\bar{v}_{it}(z_{it}^e, z_{jt}^e) = 0$.

As suggested by intuition, X_t^{FS} is the smallest threshold among the four. The ordinal ranks of X_t^{PS} , X_t^{MR} , and X_t^{LL} are affected by various DL's although X_t^{MR} never exceeds X_t^{LL} . Lemma 5.4 specifies the possibilities.

The condition $\bar{m}_{it}^S \geq \bar{m}_{jt}^S$ in Lemma 5.4 implies that the very first unit to backlog should be from class j demand, since the expected backlogging loss caused by class j demand is smaller. If $m_{jt} \leq \bar{m}_{it}^L$, the firm should keep backlogging class j demand; otherwise, start backlogging class i demand.

We introduce \bar{m}_{it}^E in Lemma 5.4 and refer to it as class i demand's Equal Dynamic Loss (EDL), since whenever c_e equals \bar{m}_{it}^E , X_t^{MR} equals X_t^{PS} , namely, partial fulfillment of both classes of regular demand occurs at the same time when capacity reservation for emergency demand is lowered to exactly $\bar{C}_{1t} \vee \bar{C}_{2t}$ units.

Note that when $X_t^{PS} = \infty$, partial fulfillment of both classes of regular demand is not allowed. There are two cases that exclude partial fulfillment of both regular demand classes: (1) One of the regular demand classes is fully satisfied and the other partially satisfied (or not satisfied at all). (2) One of the regular demand classes is partially satisfied and the other not satisfied at all. In either cases, one class of regular demand seems to dominant the other class.

As indicated by their definitions, each of the four thresholds on X_t represents a beginning or an end to a pattern of the optimal solution. The thresholds, therefore, are key to the description of the optimal policy.

5.1.3. When the firm's capacity exceeds both minimal reservation levels

As explained earlier, when the firm's capacity C exceeds $\bar{C}_{1t} \vee \bar{C}_{2t}$, the Diagonal-Stable region is not an empty set. Both regular classes of demand are processed in period t within the Diagonal-Stable region and we have $\mathbf{z}^*(\mathbf{x}_t) < \mathbf{x}_t$ for the region. That is,

$$\Omega_t^S = \{(x_{it}, x_{jt}) \geq 0 \mid x_{it} + x_{jt} = X_t, x_{it} > z_{it}^e(X_t), x_{jt} > z_{jt}^e(X_t)\}.$$

Intuitively, when the Diagonal-Stable region exists, the Only-High and Only-Low regions, if both are valid, must be on different sides of the Diagonal-Stable region. Given a fixed amount of the PL demand, the optimal policy looks like this: satisfying only the PL demand if the amount of unsatisfied PH demand is small enough (may be zero); satisfying both the PH and the PL demand if the amount of unsatisfied PH demand grows, and satisfying only the PH demand if the

amount of unsatisfied PH demand exceeds a higher threshold. The thresholds form the boundaries between the regions. Characterizing the optimal policy involves characterizing the boundaries.

For ease of presentation and without loss of generality, we will assume, for the rest of the paper, that class i of the regular demand is the PH demand and class j is the PL demand in period t , that is, $m_{it} \geq m_{jt}$.

Denote by \mathcal{H}_t and \mathcal{L}_t the borders connecting the Diagonal-Stable region to the Only-High and Only-Low regions, respectively. We have

$$\mathcal{H}_t = \{(x_{it}, x_{jt}) \geq 0 \mid x_{it} + x_{jt} = X_t, x_{it} > z_{it}^e(X_t), x_{jt} = z_{jt}^e(X_t)\},$$

$$\mathcal{L}_t = \{(x_{it}, x_{jt}) \geq 0 \mid x_{it} + x_{jt} = X_t, x_{jt} > z_{jt}^e(X_t), x_{it} = z_{it}^e(X_t)\}.$$

Note that the Only-High and Only-Low regions can be expressed mathematically as follows:

$$\Omega_t^H = \{(x_{it}, x_{jt}) \geq 0 \mid x_{it} + x_{jt} = X_t, x_{it} > z_{it}^e(X_t), x_{jt} \leq z_{jt}^e(X_t)\},$$

$$\Omega_t^L = \{(x_{it}, x_{jt}) \geq 0 \mid x_{it} + x_{jt} = X_t, x_{jt} > z_{jt}^e(X_t), x_{it} \leq z_{it}^e(X_t)\}.$$

Theorem 5.5 finds that if the Only-Low region is reduced to \mathcal{L}_t , \mathcal{L}_t must be the axis representing the amount of unsatisfied PL demand. At any state within such a line-shaped Only-Low region, there is no unsatisfied PH demand. There are two conditions under either of which the Only-Low region is reduced to \mathcal{L}_t : (i) the period types and limit types of the regular demand are consistent; or (ii) the period types and limit types of the regular demand are inconsistent, and the LDL of the PH demand is higher than the ML of the PL demand. Condition (i) is easy to grasp because when the PH demand is also the LH demand, it is apparently more costly than the PL (also LL) demand if not satisfied. In this case, if the Only-Low region does exist, it must be that there is no unsatisfied PH demand at all. Condition(ii) is a little more complicated. When the PH demand is no longer the LH demand, the priority of the PH demand is weakened. By requiring the LDL of the PH demand to be higher than the ML of the PL demand, condition (ii) strengthens the priority of the PH demand, making the reduction of the Only-Low region to \mathcal{L}_t reasonable. We state **Theorem 5.5** formally below.

Theorem 5.5. Given $C > (\bar{C}_{1t} \vee \bar{C}_{2t})$, we have $\Omega_t = \Omega_t^L + \Omega_t^S + \Omega_t^H$ and $\Omega_t^S \neq \emptyset$. Specifically, given $m_{it} \geq m_{jt}$, we have (1) $\Omega_t^L = \mathcal{L}_t \Leftrightarrow \mathcal{L}_t = \{(x_{it}, x_{jt}) \mid x_{it} = 0, x_{jt} > 0\}$; (2.1) if $m_i \geq m_j$, then $\Omega_t^L = \mathcal{L}_t$; (2.2) if $m_i < m_j$, then $\bar{m}_{it}^L \geq m_{jt} \Leftrightarrow \Omega_t^L = \mathcal{L}_t$.

When $C < (\bar{C}_{1t} \wedge \bar{C}_{2t})$, the types of the regular demand do not matter because no regular demand will be processed. When $(\bar{C}_{1t} \wedge \bar{C}_{2t}) < C \leq (\bar{C}_{1t} \vee \bar{C}_{2t})$, the period type, not limit type, matters as the firm prioritizes its demand. Specifically, the PH demand's priority over the PL demand is definitive when $(\bar{C}_{1t} \wedge \bar{C}_{2t}) < C \leq (\bar{C}_{1t} \vee \bar{C}_{2t})$. When $(\bar{C}_{1t} \vee \bar{C}_{2t}) < C$, both period type and limit type matter. We define the dominant PH demand scenario as the case in which one of the conditions for the reduction of the Only-Low region to \mathcal{L}_t holds, because under such scenario the PL demand is never fulfilled before all the PH demand is processed. We define the non-dominant PH demand scenario as the case in which neither of the aforementioned conditions is met. In order to further characterize the optimal policy, we discuss the dominant and non-dominant PH demand scenarios in **Sections 5.2** and **5.3**, respectively.

5.2. The dominant PH demand scenario

Theorem 5.5 has shown that when the PH demand is dominant, the Only-Low region is reduced to \mathcal{L}_t , which sits on the axis representing the PL demand. Our focus, in this section, is therefore on the internal structures of the Diagonal-Stable and the Only-High regions.

5.2.1. The Diagonal-Stable region

Within the Diagonal-Stable region, both the PH and PL demand are processed, at least partially. Because of the priority the PH demand

has over the LH demand here in **Section 5.2**, the fulfillment of the PL demand implies that no PH demand will be backlogged. In other words, if a state $\mathbf{x}_t = (x_{it}, x_{jt})$ belongs to the Diagonal-Stable region, the optimal policy requires that all x_{it} units of the PH demand be fulfilled in period t . Whether all x_{jt} unit of the PL demand will be fulfilled depends on how $X_t = x_{it} + x_{jt}$ compares to thresholds X_t^{FS} and X_t^{MR} .

Recall that X_t^{FS} is defined as the aggregate amount of unsatisfied regular demand below which both classes of regular demand are fully satisfied and X_t^{MR} is the aggregate amount of unsatisfied regular demand beyond which the firm makes a minimal reservation ($\bar{C}_{it} \vee \bar{C}_{jt}$) for emergency demand. With $m_{it} \geq m_{jt}$, we have $\bar{C}_{jt} = (\bar{C}_{it} \vee \bar{C}_{jt})$.

Lemma 4.3 in **Section 4.2** has shown that if the PH demand is dominant, namely, either (i) $m_i \geq m_j$ or (ii) $m_i < m_j$ and $\bar{m}_{it}^L > m_{jt}$, then $\bar{m}_{it}^S \geq \bar{m}_{jt}^S$. Satisfaction of the last condition allows us to apply **Lemma 5.4** in **Section 5.1.2**, and doing so tells us that $0 < X_t^{FS} \leq X_t^{MR} = X_t^{LL} < X_t^{PS} = \infty$. Note that X_t^{LL} is the aggregate amount of unsatisfied regular demand beyond which the desired amount of PL demand to be backlogged is linearly increasing in X_t at the rate of 1. So, $X_t^{MR} = X_t^{LL}$ suggests that as soon as the reservation for emergency demand is lowered to exactly \bar{C}_{jt} units, every additional unit of the regular demand should translate into an additional unit of the PL demand being backlogged. Also note that X_t^{PS} is the aggregate amount of unsatisfied regular demand beyond which both classes of regular demand are partially satisfied. The result that $X_t^{PS} = \infty$ is consistent with the fact that no PL demand can be processed before all the PH demand is fulfilled, which is due to the dominance of the PH demand in this section.

Corollary 5.6 formalizes the characterization of the Diagonal-Stable region.

Corollary 5.6. Suppose $C > (\bar{C}_{1t} \vee \bar{C}_{2t})$ and $m_{it} \geq m_{jt}$. Either (i) $m_i \geq m_j$ or (ii) $m_i < m_j$ and $\bar{m}_{it}^L > m_{jt}$. For any $\mathbf{x}_t \in \Omega_t^S$, $z_{it}^* = 0$, $z_{jt}^* < x_{jt}$ if $x_{jt} > 0$ and $z_{jt}^* = 0$ if $x_{jt} = 0$, and $\frac{\partial z_{jt}^*}{\partial x_{it}} = \frac{\partial z_{jt}^*}{\partial x_{jt}}$. Furthermore, 1. $X_t \leq X_t^{FS} \Rightarrow z_{jt}^* = 0$, 2. $X_t^{FS} < X_t < X_t^{MR} \Rightarrow \frac{\partial z_{jt}^*}{\partial x_{it}} \in (0, 1)$, and 3. $X_t^{MR} \leq X_t \Rightarrow z_{jt}^* = X_t - C + \bar{C}_{jt}$.

Basically, if X_t is small enough ($\leq X_t^{FS}$), all x_{jt} units of the PL demand will be satisfied and $C - (x_{it} + x_{jt}) = C - X_t$ units of capacity are reserved for emergency demand. Since $X_t^{FS} \leq X_t^{MR}$ we know that $C - X_t$ should be at least the minimal reservation capacity $\bar{C}_{jt} = \bar{C}_{it} \vee \bar{C}_{jt}$ when $X_t \leq X_t^{FS}$. If X_t is big enough ($\geq X_t^{MR}$), the firm's C units of capacity should be divided into three groups: \bar{C}_{jt} units will be reserved for emergency demand, x_{it} units will be used to fulfill all the PH demand, and the rest $(x_{jt} - z_{jt}^*)$ units of capacity will be used to fulfill the PL demand where z_{jt}^* is $\bar{C}_{jt} + x_{it} + x_{jt} - C = X_t - (C - \bar{C}_{jt})$, the difference between the aggregate amount of unsatisfied regular demand X_t and the firm's capacity after the minimal reservation \bar{C}_{jt} . So, when $X_t \geq X_t^{MR}$, every additional unit of the regular demand translates into an additional unit of the PL demand being backlogged. Interestingly, since z_{jt}^* , the optimal amount of the PL demand to be backlogged in period t , is determined by X_t , as opposed to x_{jt} , z_{jt}^* is affected by the amount of unsatisfied PH and PL demand equally, instead of by the amount of unsatisfied PL demand alone. If X_t is between X_t^{FS} and X_t^{MR} , z_{jt}^* is between zero and $X_t - (C - \bar{C}_{jt})$. Furthermore, an additional unit of regular demand does not necessarily result in an increase in z_{jt}^* because the capacity reservation for emergency demand can be lowered from $C - X_t^{FS}$ toward \bar{C}_{jt} when $X_t^{FS} < X_t < X_t^{MR}$.

It is interesting to note that for any state in the Diagonal-Stable region that corresponds to a given X_t , the optimal amount of the PH demand to be backlogged is zero and the optimal amount of the PL demand to be backlogged is a constant. Recall that \mathcal{H}_t is the Diagonal-Stable region's border with the Only-High region. Let $\mathcal{H}_t(X_t)$ represent the intersection of \mathcal{H}_t and the diagonal line that contains

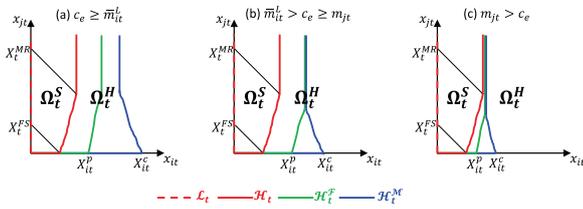


Fig. 1. The segmentation of Ω_t under the optimal policy when $C > (\bar{C}_{1t} \vee \bar{C}_{2t})$, $m_{it} \geq m_{jt}$ and $m_i \geq m_j$.

all the states with the same aggregate amount of unsatisfied regular demand: X_t . The optimal amount of the PL demand to be backlogged given X_t is just the coordinate of $\mathcal{H}_t(X_t)$ that represents the PL demand.

5.2.2. The Only-High region

For the Only-High region, our focus is on the optimal amount of the PH demand to be backlogged z_{it}^* (\mathbf{x}_t).

When the horizontal axis represents the unsatisfied PH demand, the Only-High region is to the right of the Diagonal-Stable region. Recall that \mathcal{H}_t is the border between the Only-High region and the Diagonal-Stable region. That is, all points inside the Only-High region are to the right of \mathcal{H}_t . In order to describe how z_{it}^* changes within the Only-High region, we further define two switching curves: \mathcal{H}_t^F and \mathcal{H}_t^M . \mathcal{H}_t^F is closer to \mathcal{H}_t than \mathcal{H}_t^M is. If \mathbf{x}_t is between \mathcal{H}_t and \mathcal{H}_t^F , then $z_{it}^* = 0$ (i.e. all the unsatisfied PH demand will be fulfilled in period t) and at least \bar{C}_{it} units of the firm’s capacity are reserved for emergency demand. If \mathbf{x}_t is between \mathcal{H}_t^F and \mathcal{H}_t^M , then some PH demand will be backlogged and at least \bar{C}_{it} units of the firm’s capacity are reserved for emergency demand. If \mathbf{x}_t is to the right of \mathcal{H}_t^M , then $z_{it}^* = x_{it} - (C - \bar{C}_{it})$ (i.e. $C - \bar{C}_{it}$ units of the PH demand will be fulfilled) and exactly \bar{C}_{it} units of capacity are reserved for emergency demand.

Recall that we use the thresholds X_{it}^P and X_{it}^C to separate the Only-High region into three parts when $(\bar{C}_{1t} \wedge \bar{C}_{2t}) \leq C \leq (\bar{C}_{1t} \vee \bar{C}_{2t})$ (in Section 5.1). If we were to draw \mathcal{H}_t^F and \mathcal{H}_t^M in that case, \mathcal{H}_t^F would be a straight line that parallels the vertical axis (representing the PL demand) and intersects with the horizontal axis (representing the PH demand) at $x_{it} = X_{it}^P$, and \mathcal{H}_t^M would also parallel the vertical axis and intersect the horizontal axis at $x_{it} = X_{it}^C$. When $(\bar{C}_{1t} \vee \bar{C}_{2t}) < C$, \mathcal{H}_t^F still intersects with the horizontal axis at $x_{it} = X_{it}^P$ and \mathcal{H}_t^M still intersects with the horizontal axis at $x_{it} = X_{it}^C$. But, neither \mathcal{H}_t^F nor \mathcal{H}_t^M is a straight line when $(\bar{C}_{1t} \vee \bar{C}_{2t}) < C$. That is, the level of the unsatisfied PL demand (i.e. x_{jt}) affects the thresholds on x_{it} here.

We characterize the shapes and positions of \mathcal{H}_t^F , \mathcal{H}_t^M and \mathcal{H}_t in Lemma A.8 in Appendix A.4.

We create Fig. 1 in order to show how the state space Ω_t is mapped into the Only-High and Diagonal-Stable regions when $C > (\bar{C}_{1t} \vee \bar{C}_{2t})$, $m_{it} \geq m_{jt}$, and $m_i \geq m_j$. Note that the Full-Reservation region does not exist and the Only-Low region is reduced to the vertical border of the first quadrant under these conditions. While the shapes and sizes of the three subregions of Ω_t^S remain the same as c_e changes, the shapes and sizes of the three subregions of Ω_t^H depend on which range c_e falls into.

We make two observations from Fig. 1. Note that the horizontal axis represents the PH demand.

- The intersection of \mathcal{H}_t and the horizontal axis is the state at which there are X_{it}^{FS} units of unsatisfied PH demand and zero unsatisfied PL demand. Given the rightward leaning \mathcal{H}_t , any state with fewer units of unsatisfied PH demand than X_{it}^{FS} must be in the Diagonal-Stable region, as opposed to the Only-High region.
- Although both \mathcal{H}_t and \mathcal{H}_t^F are initially rightward leaning and \mathcal{H}_t^M is initially leftward leaning as they start from the horizontal axis,

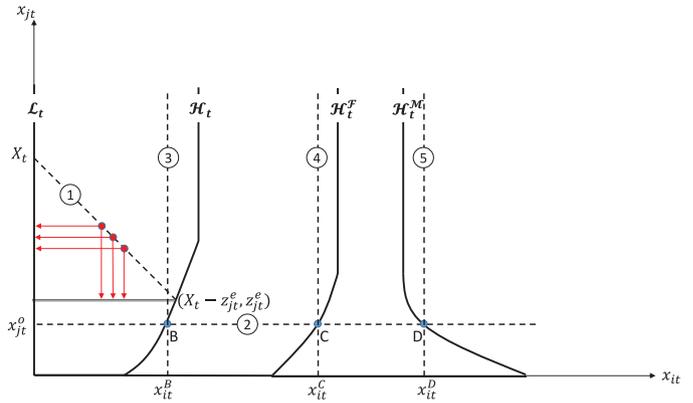


Fig. 2. The position of five representative lines when $C > (\bar{C}_{1t} \vee \bar{C}_{2t})$, $m_{it} \geq m_{jt}$ and $m_i \geq m_j$.

all three curves become straight lines when the amount of unsatisfied PL demand x_{jt} is big enough. That is, changes in x_{jt} no longer affect the positions of the curves when x_{jt} is big enough.

5.2.3. Graphs of the optimal policy

In the above two sections, we explain the structure of the optimal policy by describing the borders outlining the Diagonal-Stable and Only-High regions and the thresholds within the two regions. Based on the propositions and lemmas we provide in those sections, we are also able to plot graphs depicting the optimal policy directly. In Figs. 2–4, we present graphical illustrations of the optimal policy from various perspectives given that $C > (\bar{C}_{1t} \vee \bar{C}_{2t})$, $m_{it} \geq m_{jt}$ and $m_i \geq m_j$. In these figures, whenever the dominant PH demand (Class i) is plotted, it is represented by the horizontal axis, and whenever the PL demand (Class j) is plotted, it is represented by the vertical axis.

Our state-space is two-dimensional and so is the space for the optimal policy. Instead of attempting to plot four-dimensional figures, we show what the optimal policy looks like from the perspectives of five lines in two-dimensional figures. Fig. 2 depicts the positions of the five representative lines in the state space Ω_t . Line (1) is representative of any 135 degree line within the Diagonal-Stable region. By moving Line (1) parallelly within the Diagonal-Stable region, we can cover the whole region. Line (2) is parallel to the horizontal axis. As Line (2) moves upward and downward, it can cover the whole state space. The purpose of Line (2) is to illustrate how z_{it}^* and z_{jt}^* change with respect to x_{it} when x_{jt} is fixed. Lines (3), (4) and (5) are parallel to the vertical axis. They intersect with \mathcal{H}_t , \mathcal{H}_t^F and \mathcal{H}_t^M , respectively. We outline the look of the optimal policy as x_{jt} changes when x_{it} is fixed via Lines (3), (4) and (5).

Fig. 2 also describes how to find the optimal policy for any state on Line (1). Line (1) is a 135 degree line between the vertical axis and \mathcal{H}_t , i.e. the border between the Diagonal-Stable region and the Only-High region. For each state on the 135 degree line, the corresponding optimal policy (z_{it}^*, z_{jt}^*) equals $(0, z_{jt}^e)$. Recall that we can find (z_{it}^e, z_{jt}^e) for each X_t , and all states on the same 135 degree line have the same X_t . For each 135 degree line, the corresponding z_{jt}^e is just the vertical coordinate of the intersection of the line and \mathcal{H}_t . Note that the length of the horizontal arrow originated from a state on Line (1) represents the amount of PH demand that will be satisfied within period t and the length of the vertical arrow originated from a state on Line (1) represents the amount of PL demand that will be satisfied within period t . Clearly, all unsatisfied PH demand (x_{it}) will be processed within period t .

As shown in Fig. 2, Line (2) intersects with \mathcal{H}_t , \mathcal{H}_t^F , and \mathcal{H}_t^M at Points B, C, and D, respectively. We plot z_{it}^* and z_{jt}^* as functions of x_{it} separately in Fig. 3. In both plots, x_{jt} is fixed at x_{jt}^0 , which is the vertical coordinate of Points B, C, and D. z_{it}^* starts from zero when x_{it} is zero.

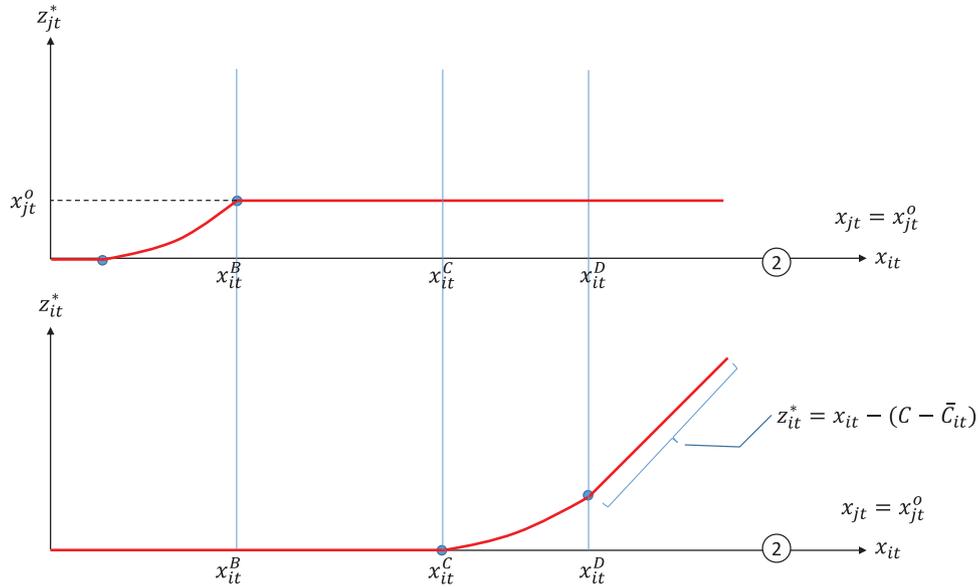


Fig. 3. The optimal policy corresponding to Line (2) when $C > (\bar{c}_{1t} \vee \bar{c}_{2t})$, $m_{it} \geq m_{jt}$ and $m_i \geq m_j$.

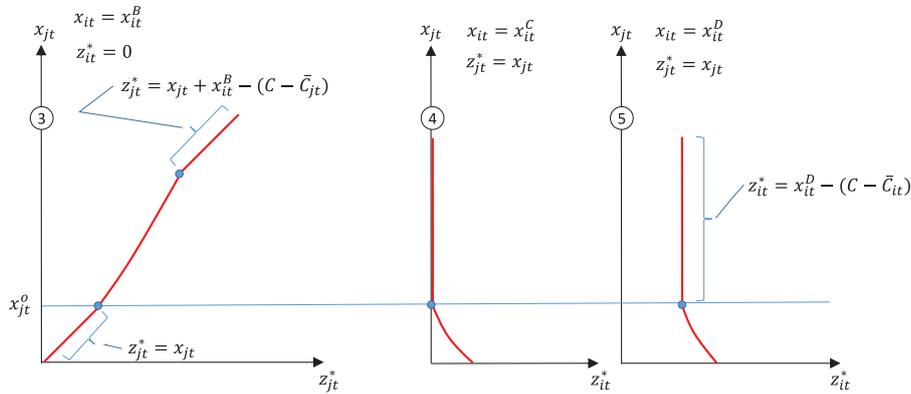


Fig. 4. The optimal policy corresponding to Lines (3)–(5) when $C > (\bar{c}_{1t} \vee \bar{c}_{2t})$, $m_{it} \geq m_{jt}$ and $m_i \geq m_j$.

As x_{it} increases, z_{it}^* remains at zero until x_{it} reaches x_{it}^C , which is the horizontal coordinate of Point C. z_{it}^* increases strictly, thereafter. z_{it}^* 's rate of increase is smaller than one before x_{it} reaches x_{it}^D , which is the horizontal coordinate of Point D, becomes one when x_{it} reaches x_{it}^D , and stays at one as x_{it} continues to increase. Note that $z_{it}^* = x_{it} - (C - \bar{c}_{it})$ when x_{it} exceeds x_{it}^D . z_{jt}^* also starts from zero when x_{it} is zero. As x_{it} increases, z_{jt}^* stays at zero initially and then increases strictly until x_{it} reaches x_{it}^B , which is the horizontal coordinate of Point B. At that state, which is (x_{it}^B, x_{jt}^0) , z_{jt}^* equals x_{jt}^0 . As x_{it} continues to increase, z_{jt}^* stays at x_{jt}^0 . Note that Line (2) enters the Only-High region once it crosses Point B from the Diagonal-Stable region.

Line (3) is parallel to the vertical axis and intersects \mathcal{H}_t (the border between the Diagonal-Stable and the Only-High regions) at Point B. The higher part of Line (3) is inside the Diagonal-Stable region and the lower part is inside the Only-High region. For any state on Line (3), $z_{it}^* = 0$. In the first plot of Fig. 4, we fix x_{it} at x_{it}^B , i.e. the horizontal coordinate of Point B, and plot how z_{jt}^* changes as x_{jt} increases. We write (3) on the vertical axis of this plot to distinguish it from others. z_{jt}^* increase strictly as x_{jt} increases. Initially, $z_{jt}^* = x_{jt}$. That is, z_{jt}^* 's rate of increase is one. The rate drops below one after x_{jt} reaches x_{jt}^0 . The smaller-than-one rate lasts for a while as x_{jt} increases. Eventually, the rate of increase goes up to one. Once there, $z_{jt}^* = x_{jt} + x_{it}^B - (C - \bar{c}_{jt})$ and z_{jt}^* 's rate of increase stays at one as x_{jt} continues to increase. That is, for states lower than Point B on Line (3), z_{jt}^* increases at the same

rate as x_{jt} ; for states higher than Point B on Line (3), z_{jt}^* increases slower than x_{jt} initially as a result of going from the Only-High region to the Diagonal-Stable region, but z_{jt}^* 's rate of increase eventually goes up to one.

Lines (4) and (5) are parallel to the vertical axis and intersect \mathcal{H}_t^F and \mathcal{H}_t^M at Points C and D, respectively. Lines (4) and (5) are inside the Only-High region because the thresholds \mathcal{H}_t^F and \mathcal{H}_t^M are within the Only-High region. The definition of the Only-High region implies that for any state on Lines (4) and (5), z_{jt}^* equals x_{jt} . For Lines (4) and (5), we fix x_{it} at x_{it}^C and x_{it}^D , respectively, and plot z_{it}^* as x_{jt} changes in Fig. 4. We write (4) on the vertical axis in the plot for Line (4) and do the same for Line (5). For Line (4), as x_{jt} increases, z_{it}^* decreases strictly until x_{jt} reaches x_{jt}^0 , which is the vertical coordinate of Points B, C and D. z_{it}^* drops to 0 when $x_{jt} = x_{jt}^0$ and stays at 0 as x_{jt} further increases. The optimal policy for Line (5) is similar but slightly different. For Line (5), as x_{jt} increases, z_{it}^* also decreases strictly until x_{jt} reaches x_{jt}^0 . However, z_{it}^* is strictly positive when $x_{jt} = x_{jt}^0$. At the state (x_{it}^D, x_{jt}^0) , $z_{it}^* = x_{it}^D - (C - \bar{c}_{it})$. As x_{jt} further increases, z_{it}^* remains the same.

5.3. The non-dominant PH demand scenario

When $t \leq T^0$, the regular demand's period and limit types are inconsistent. Given our assumption of $m_{it} \geq m_{jt}$, we have $m_i < m_j$ in period $t \leq T^0$. In addition to the inconsistency of period and limit types, we also require $m_{jt} \leq \bar{m}_{it}^L$ to hold for the PH demand to be

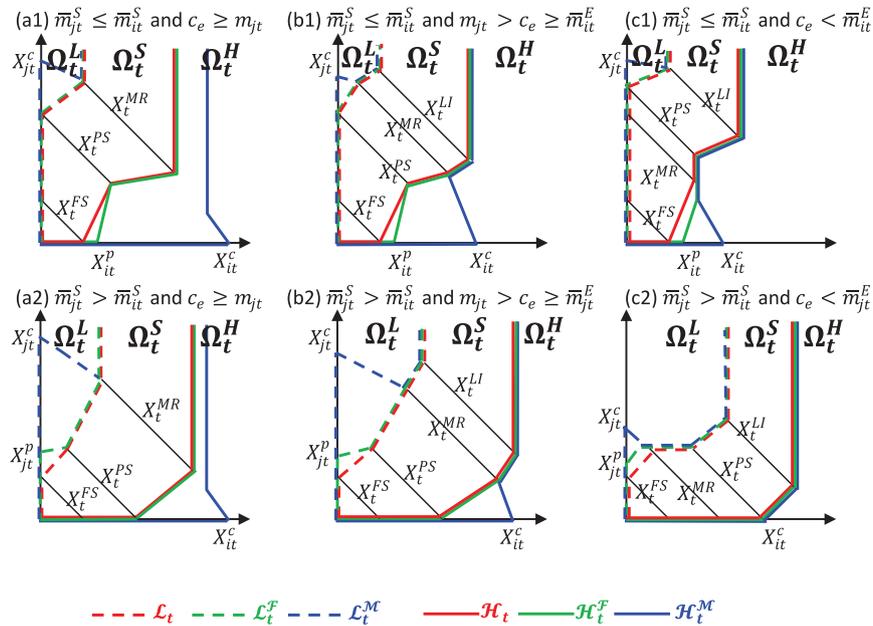


Fig. 5. The segmentation of Ω_t under optimal policy when $C > (\bar{C}_{it} \vee \bar{C}_{2t})$, $m_{it} \geq m_{jt}$, $m_i < m_j$, and $\bar{m}_{it}^L < m_{jt}$.

non-dominant. We examine what happens under these conditions in this section. As in the previous section, we use the horizontal axis to represent the PH demand in our graphical descriptions.

5.3.1. The Only-High region

The analysis on the Only-High region when the PH demand is non-dominant is similar to that in Section 5.2. Again, the characterizations of the switching curves \mathcal{H}_t^F and \mathcal{H}_t^M are at the core of the analysis. The shapes of \mathcal{H}_t^F and \mathcal{H}_t^M , however, are more varied as shown in Fig. 5. The non-dominance of the PH demand implies that more parameter comparisons are needed when determining the optimal policy.

When the PH demand is dominant, we need to know where c_e lands with respect to the interval $[m_{jt}, \bar{m}_{it}^S]$. When the PH demand is non-dominant, we need to know where c_e lands with respect to either the interval $[\bar{m}_{it}^E, m_{jt}]$ or the interval $[\bar{m}_{jt}^E, m_{jt}]$, in addition to the ordinal ranks of \bar{m}_{it}^S and \bar{m}_{jt}^S . Lemmas A.9 and Appendix A.10 in Appendix A.4 provide the details.

5.3.2. The Only-Low region

In the Only-Low region, the PH demand is not satisfied at all ($z_{it}^* = x_{it}$) and the amount of the PL demand to be backlogged is $z_{jt}^* = \bar{z}_{jt}(x_{it}, X_t)$ in period t .

As in the Only-High region, we can also identify two switching curves \mathcal{L}_t^F and \mathcal{L}_t^M here, splitting the Only-Low region into three areas with unique patterns of z_{jt}^* . Since the horizontal axis represents the PH demand, the Only-Low region is to the left of the Diagonal-Stable region with the border \mathcal{L}_t between them. \mathcal{L}_t^F is closer to \mathcal{L}_t than \mathcal{L}_t^M is. If the state \mathbf{x}_t is between \mathcal{L}_t and \mathcal{L}_t^F , then $z_{jt}^* = 0$ (i.e. all the PL demand will be fulfilled). If \mathbf{x}_t is between \mathcal{L}_t^F and \mathcal{L}_t^M , then some PL demand will remain unfulfilled and at least \bar{C}_{jt} units of capacity are reserved for emergency demand. If \mathbf{x}_t is to the left of \mathcal{L}_t^M , then $z_{jt}^* = x_{jt} - (C - \bar{C}_{jt})$ and exactly \bar{C}_{jt} units of capacity are reserved for emergency demand. Lemmas A.11 and Appendix A.12 in Appendix A.4 describe the shapes and positions of \mathcal{L}_t^F and \mathcal{L}_t^M under various parameter combinations.

Although the two regions share structural similarities, the Only-High and Only-Low regions are not symmetrical. The Only-high region has a lower bound on the PH demand and no upper bound on the PL demand while the Only-Low region has a lower bound on the PL demand and also an upper bound on the PH demand. That is, for

the PH (or the PL demand) to be the sole type of regular demand that gets any capacity allocation, the amount of unsatisfied PH (or PL) demand must exceed certain threshold. Meanwhile, if there is too much unsatisfied PH demand, the firm will have to allocate some capacity to it. However, it is possible for the firm to fulfill only the PH demand even if the amount of unsatisfied PL demand goes to infinity.

5.3.3. The Diagonal-Stable region

Inside the Diagonal-Stable region, both classes of the regular demand are processed. That is, $\mathbf{z}_t^* < \mathbf{x}_t$. Applying the definition of \mathbf{z}_t^* (from Section 4.4), we know that $\mathbf{z}_t^* = \mathbf{z}_t^*(X_t) < \mathbf{x}_t$ for each state inside the Diagonal-Stable region. The optimal amount of demand to be backlogged (\mathbf{z}_t^*) is therefore affected by the state \mathbf{x}_t through the aggregate amount of unsatisfied regular demand beyond which both classes of regular demand are partially satisfied. These four thresholds: X_t^{FS} , X_t^{PS} , X_t^{MR} , and X_t^{LI} , shape the turning points of the optimal policy and segment different patterns of the optimal policy in the Diagonal-Stable region in the non-dominant PH demand scenario. The ordinal ranks of the four thresholds are discussed in Lemma 5.4 in Section 5.1.2. The possibilities listed there shed light on the characterization of the optimal policy here.

Recall that we need the thresholds X_t^{FS} and $X_t^{MR} = X_t^{LI}$ on X_t when discussing the Diagonal-Stable region in the dominant PH demand scenario. When the PH demand is non-dominant, X_t^{LI} is no longer guaranteed to equal X_t^{MR} and X_t^{PS} is no longer infinity. Recall that X_t^{PS} is the aggregate amount of unsatisfied regular demand beyond which both classes of regular demand are partially satisfied. These four thresholds: X_t^{FS} , X_t^{PS} , X_t^{MR} , and X_t^{LI} , shape the turning points of the optimal policy and segment different patterns of the optimal policy in the Diagonal-Stable region in the non-dominant PH demand scenario. The ordinal ranks of the four thresholds are discussed in Lemma 5.4 in Section 5.1.2. The possibilities listed there shed light on the characterization of the optimal policy here.

X_t must be in one of these three intervals: $[0, X_t^{FS}]$, (X_t^{FS}, X_t^{LI}) , and $[X_t^{LI}, +\infty)$. Let $B = \mathcal{H}, \mathcal{L}$. We use $B_t(X_t)$ to denote a point on B_t such that the aggregate amount of unsatisfied regular demand is X_t at the state. The coordinates of $B_t(X_t)$ are $B_{it}(X_t)$ and $B_{jt}(X_t)$.

- When $X_t \leq X_t^{FS}$, it is clear that $\mathcal{L}_t(X_t) = (0, X_t)$ and $\mathcal{H}_t(X_t) = (X_t, 0)$.
- When $X_t \geq X_t^{LI}$, by the definition of X_t^{MR} and X_t^{LI} , we know $\frac{\partial \mathcal{H}_{it}}{\partial X_t} = \frac{\partial \mathcal{L}_{it}}{\partial X_t} = 0$, $\frac{\partial \mathcal{H}_{jt}}{\partial X_t} = \frac{\partial \mathcal{L}_{jt}}{\partial X_t} = 1$, and $\mathcal{H}_{it} - \mathcal{L}_{it} = C - \bar{C}_{jt}$.
- When $X_t^{FS} < X_t < X_t^{LI}$, $B_{it}(X_t) = X_t - B_{jt}(X_t)$ for $B = \mathcal{H}, \mathcal{L}$. Under this condition, Lemma 5.7 characterizes the shapes and positions of \mathcal{L}_t and \mathcal{H}_t , making use of the thresholds X_t^{PS} and X_t^{MR} .

Table 1
Overview of the optimal policy.

Regions	Existence conditions	Thresholds or switching curves
Full-Reservation	$C \leq (\bar{C}_{it} \wedge \bar{C}_{jt})$	N/A
Only-High	$(\bar{C}_{it} \wedge \bar{C}_{jt}) < C \leq (\bar{C}_{it} \vee \bar{C}_{jt})$	X_{it}^P and X_{it}^C
	$(\bar{C}_{it} \vee \bar{C}_{jt}) < C$	\mathcal{H}_{it}^F (meets the PH demand axis at X_{it}^P) \mathcal{H}_{it}^M (meets the PH demand axis at X_{it}^C)
Diagonal-Stable	$(\bar{C}_{it} \vee \bar{C}_{jt}) < C$	X_{it}^{FS} , X_{it}^{PS} , X_{it}^{MR} , and X_{it}^{LI}
Only-Low	$(\bar{C}_{it} \vee \bar{C}_{jt}) < C$	\mathcal{L}_{it}^F (meets the PL demand axis at X_{it}^P)
		\mathcal{L}_{it}^M (meets the PL demand axis at X_{it}^C)

Lemma 5.7. Let $p = i$ if $\bar{m}_{jt}^S \leq \bar{m}_{it}^S$ and $p = j$ otherwise. Consider $C > (\bar{C}_{it} \vee \bar{C}_{jt})$. Given $m_{it} \geq m_{jt}$, $m_i < m_j$, $\bar{m}_{it}^L < m_{jt}$, and $X_{it}^{FS} < X_t < X_{it}^{LI}$, we have

- If $c_e \geq m_{jt}$, (i) $X_t \leq X_t^{PS} \Rightarrow \frac{\partial \mathcal{I}_{pt}}{\partial X_t} \in (0, 1)$ and $\mathcal{J}_{pt} = 0$; (ii) $X_t^{PS} \leq X_t < X_t^{LI} \Rightarrow \frac{\partial \mathcal{I}_{pt}}{\partial X_t}, \frac{\partial \mathcal{J}_{pt}}{\partial X_t} \in (0, 1)$.
- If $m_{jt} > c_e \geq \bar{m}_{pt}^E$, (i) $X_t \leq X_t^{PS} \Rightarrow \frac{\partial \mathcal{I}_{pt}}{\partial X_t} \in (0, 1)$ and $\mathcal{J}_{pt} = 0$; (ii) $X_t^{PS} \leq X_t < X_t^{MR} \Rightarrow \frac{\partial \mathcal{I}_{pt}}{\partial X_t}, \frac{\partial \mathcal{J}_{pt}}{\partial X_t} \in (0, 1)$; (iii) $X_t^{MR} \leq X_t < X_t^{LI} \Rightarrow \mathcal{I}_{pt} - \mathcal{J}_{pt} = C - \bar{C}_{jt}$ and $\frac{\partial \mathcal{I}_{pt}}{\partial X_t} \in (0, 1)$.
- If $c_e < \bar{m}_{pt}^E$, (i) $X_t \leq X_t^{MR} \Rightarrow \frac{\partial \mathcal{I}_{pt}}{\partial X_t} \in (0, 1)$ and $\mathcal{J}_{pt} = 0$; (ii) $X_t^{MR} \leq X_t < X_t^{PS} \Rightarrow \mathcal{I}_{pt} = C$ and $\mathcal{J}_{pt} = 0$; (iii) $X_t^{PS} \leq X_t < X_t^{LI} \Rightarrow \mathcal{I}_{pt} - \mathcal{J}_{pt} = C$ and $\frac{\partial \mathcal{I}_{pt}}{\partial X_t} \in (0, 1)$.

The borders and internal structures of the Diagonal-Stable, Only-Low and Only-High regions are complicated when the PH demand is not dominant. We create Fig. 5 in order to illustrate the segmentation of the state space Ω_t when $C > (\bar{C}_{it} \vee \bar{C}_{jt})$, $m_{it} \geq m_{jt}$, $m_i < m_j$, and $\bar{m}_{it}^L < m_{jt}$. Depending on which range c_e falls into and how \bar{m}_{it}^S compares to \bar{m}_{jt}^S , the borders and internal structures of the three regions have six different looks.

In Fig. 5, cases (a1), (b1), and (c1) correspond to the condition $\bar{m}_{it}^S \geq \bar{m}_{jt}^S$, and cases (a2), (b2), and (c2) are under the condition $\bar{m}_{it}^S < \bar{m}_{jt}^S$. Since the first unit of backlogged demand should be from the class of regular demand with a smaller SDL, the Only-Low region becomes larger when the PL demand's (class j 's) SDL exceeds the PH demand's.

5.4. Summary of the optimal policy

The optimal policy is characterized by various thresholds and switching curves. Table 1 provides a summary of the thresholds and switching curves, which segment the regions. Each segmentation features a unique pattern of optimal actions.

The optimal policy involves allocating the firm's capacity to emergency demand and two classes of regular demand. We summarize, below, the behaviors of the optimal policy from the perspectives of reservation, priority assignment and competition for capacity.

Reservation property. When the PL demand is not allowed to access capacity, either $\bar{C}_{it} = \bar{C}_{it} \wedge \bar{C}_{jt}$ (i.e. the minimal reservation driven by the PH demand) or more units of capacity are reserved for emergency demand. When the PL demand has a claim on capacity, either $\bar{C}_{jt} = \bar{C}_{it} \vee \bar{C}_{jt}$ (i.e. the minimal reservation driven by the PL demand) or more units of capacity are reserved for emergency demand.

As long as $C > (\bar{C}_{it} \wedge \bar{C}_{jt})$, there is a threshold or switching curve beyond which the optimal reservation for emergency demand is exactly \bar{C}_{it} or \bar{C}_{jt} units. There is also a threshold or switching

curve below which the firm reserves at least \bar{C}_{it} or \bar{C}_{jt} units of capacity for emergency demand in addition to satisfying all regular demand.

Priority property. We identify conditions under which the PH demand is dominant. When the PH demand is dominant, the PL demand is in contention for capacity allocation only after all the PH demand is fulfilled. When the PH demand is non-dominant (i.e. $m_{it} \geq m_{jt}$, $m_i < m_j$, and $\bar{m}_{it}^L < m_{jt}$), anything can happen: both the PH demand and the PL demand could be the only class of regular demand that is fulfilled; both the PH demand and the PL demand can be partially fulfilled; either class of regular demand can be completely fulfilled while the other class is partially fulfilled.

Competition property. Naturally, the more unsatisfied demand there is for one class of regular demand, the more capacity will be allocated to this class (in the weak sense). This is true whether the PH demand is dominant or not.

When the PH demand is dominant, the capacity allocated to the PH demand never decreases, and may increase even as the amount of unsatisfied PL demand grows, while the capacity allocated to the PL demand never increases as the amount of unsatisfied PH demand grows. Corollary 5.8 formalizes this result.

Corollary 5.8. Let $C > (\bar{C}_{it} \vee \bar{C}_{jt})$ and $m_{it} \geq m_{jt}$. Suppose either (i) $m_i \geq m_j$ or (ii) $m_i < m_j$ and $\bar{m}_{it}^L \geq m_{jt}$ holds. We have $y_{it}^*(\mathbf{x}_t) = x_{it} - z_{it}^*(\mathbf{x}_t)$ with $\frac{\partial y_{it}^*}{\partial x_{it}} \in [0, 1]$, $\frac{\partial y_{it}^*}{\partial x_{jt}} \in [0, 1]$, $\frac{\partial y_{jt}^*}{\partial x_{jt}} \in [0, 1]$ and $\frac{\partial y_{jt}^*}{\partial x_{it}} \in [-1, 0]$.

When the PH demand is non-dominant, the amount of capacity allocated to one class of regular demand does not always respond to changes in the competing class's unsatisfied demand in a monotonic way. As the amount of unsatisfied demand in one class grows, the capacity allocated to the competing class may decrease, increase, or first increase and then decrease. Several factors influence the trend, for example, c_e , EDL's, SDL's and the state \mathbf{x}_t . Corollary 5.9 provides the details.

Corollary 5.9. Let $C > (\bar{C}_{it} \vee \bar{C}_{jt})$ and $m_{it} \geq m_{jt}$. Given $m_i < m_j$ and $\bar{m}_{it}^L < m_{jt}$, we have $y_{it}^*(\mathbf{x}_t) = x_{it} - z_{it}^*(\mathbf{x}_t)$ with $\frac{\partial y_{it}^*}{\partial x_{it}} \in [0, 1]$ and $|\frac{\partial y_{it}^*}{\partial x_{jt}}| \in [0, 1]$. More specifically,

- Given $\bar{m}_{it}^S \geq \bar{m}_{jt}^S$, if $c_e < \bar{m}_{it}^E$, then y_{jt}^* decreases in x_{it} ; otherwise, there exist $\bar{x}_{jt} < \bar{x}_{jt}$ so that (i) if $x_{jt} < x_{jt} < \bar{x}_{jt}$, y_{jt}^* first increases then decreases in x_{it} ; (ii) otherwise, y_{jt}^* decreases in x_{it} .
- Given $\bar{m}_{it}^S < \bar{m}_{jt}^S$, if $c_e < \bar{m}_{it}^E$, then y_{it}^* decreases in x_{jt} ; otherwise, there exist $\bar{x}_{it} < \bar{x}_{it}$ so that (i) $x_{it} \leq x_{it} \Rightarrow y_{it}^*$ decreases in x_{jt} ; (ii) $x_{it} < x_{it} < \bar{x}_{it} \Rightarrow y_{it}^*$ first increases then decreases in x_{jt} ; (iii) if $\bar{x}_{it} < x_{it}$, then y_{it}^* is monotonic in x_{jt} , and $c_e \geq m_{jt} \Leftrightarrow y_{it}^*$ increases in x_{jt} .

6. Numerical studies

The optimal policy we have characterized is complex. Its implementation requires carefully written computer programs. We design our numerical studies with two purposes in mind. First, we would like to use a numerical example to further illustrate how an optimal policy looks. Second, practitioners would be interested to know if there are simple heuristics that work well. We evaluate, here, the performance of a few heuristics in this section. Some of them reflect current industrial practices and others are designed based on our derivation of the optimal policy.

The heuristics we examine here take the following steps to solve the firm's reservation and prioritization problem: First, reserve capacity for emergency orders. Second, assign all the remaining capacity or the amount needed, whichever is smaller, to the class of regular orders with a higher priority. Third, if there is still capacity left, satisfy the class of regular orders with a lower priority as many as possible.

The three steps are implemented at the beginning of each period. The heuristics differ from each other in how the reservation quantity is determined, and/or which class of regular orders is assigned a higher priority.

The prioritization policies. For the prioritization problem, we examine two heuristics: *OptR-LH1st* and *OptR-PH1st*. *OptR* refers to the practice of choosing the optimal reservation quantity based on an exhaustive search. While *LH1st* always assigns a higher priority to the class of orders determined to be LH, *PH1st* assigns a higher priority to the class of orders determined to be PH in the current period. If the planning horizon is long enough ($> T^0$), PH orders become LH orders in any period $t > T^0$. In order to showcase the impact of the prioritization policies *LH1st* and *PH1st*, we design a numerical study in which the length of a planning horizon does not exceed T^0 . Specifically, the following parameters are considered in Numerical Study 1.

$\alpha = 0.98$; $x_{1T} = x_{2T} = 5$; $r_1 = 15$; $b_1 = 2$; $r_2 = 2$; $b_2 = 3$;
 $C = 5, 10, 15, 20, 25$; $c_e = 20, 30, 40, 50$; $T = 6, 8, 10, 12$.

Each class of regular orders and the emergency orders have the same demand distribution, which can be one of the following: (1) uniform distribution with support $[1, 10]$; (2) triangle distribution with mean = 5.5 and coefficient of variation = 0.1, 0.2, 0.3, 0.4.

We do not observe any structural change induced by probabilistic distribution parameters. The values of the thresholds are, naturally, different with different distribution parameters. We make the following observations from Numerical Study 1.

□ *PH1st* is a much better prioritization policy than *LH1st*. Under each of the parameter combinations we have tested, *PH1st* dominates *LH1st*. On average, *OptR-PH1st* generates 53.61 percent more expected profit than *OptR-LH1st* does.

Static prioritization policies that are similar to ABC classification are common in industry practices. *LH1st* is the most reasonable static prioritization policy we can think of. Nevertheless, Numerical Study 1 shows that *LH1st* significantly underperforms *PH1st*, a dynamic prioritization policy that reevaluates the regular orders' priorities each period.

□ When combined with the reservation policy *OptR*, *PH1st*'s performance is very close to that of the optimal policy. On average, the heuristic *OptR-PH1st* generates 0.66 percent less expected profit than the optimal policy does.

PH1st adjusts in each period the regular orders' priorities based on accumulated backlogging costs from the current period onward. This is also how the optimal policy determines the regular orders' priority. While the optimal policy may process PL orders before all PH orders are satisfied, *PH1st* does not fulfill any PL orders until all PH orders are satisfied. Since PL orders are not as profitable, not processing enough PL orders does not hurt the performance of *PH1st* much.

□ At extreme capacity levels, the benefit of applying the optimal policy is not evident. Fig. 6(a) illustrates this point.

When there is ample capacity, heuristics do not hurt the firm much because most of the orders can be processed without penalty regardless of the reservation and prioritization policies. When capacity is scarce, the optimal policy does not improve the firm's profit much because the firm has to let most regular orders wait and/or lose emergency orders anyway. When the firm's capacity is at intermediate levels, some orders are satisfied while others have to wait or be lost. In this case, choosing the class and amount of orders to fulfill make a difference because orders belonging to different classes have different financial consequences.

The reservation policies. We consider three simple reservation policies: *NoR*, *AvgR* and *MinR*. *NoR* refers to the practice of not making any reservation for emergency orders. From our conversations with practitioners, we have learned that although the necessity of reserving capacity for emergency orders is recognized, capacity

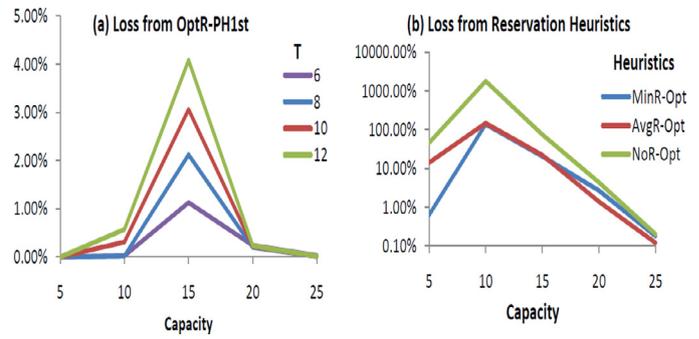


Fig. 6. The percentage loss from using heuristics versus the optimal policy (logarithm scaled).

reservations are rare in practice for lack of guidance. *AvgR* refers to the practice of reserving $\frac{\bar{C}_{1t} + \bar{C}_{2t}}{2}$ units of capacity for emergency orders. In *MinR*, the amount of capacity reserved for emergency orders depends on how x_{ht} compares to $C - \bar{C}_{lt}$. If $x_{ht} \leq C - \bar{C}_{lt}$ then \bar{C}_{lt} units of capacity are reserved for emergency orders. If $x_{ht} > C - \bar{C}_{lt}$ then \bar{C}_{ht} units of capacity is reserved for emergency orders. Note that the subscript h refers to either Class 1 or 2, whichever is determined to be the PH type. Similarly, the subscript l refers to either Class 1 or 2, whichever is determined to be the PL type.

Matched with the prioritization policy *PH1st*, we evaluate three heuristics in Numerical Study 2. All the parameters except b_2 take the same values here as those in Numerical Study 1. We set b_2 to 1 here. As a result, the regular orders have consistent types and *PH1st* becomes the optimal prioritization policy. The performances of the heuristics now rest on the reservation policies of interest.

We make the following observations from Numerical Study 2.

□ A firm's reservation policy has a significant impact on its expected profit. As shown in Fig. 6(b), the consequence of not making any reservation can be devastating. Simple reservation policies like *AvgR* and *MinR* definitely improve the situation but neither is guaranteed to generate expected profits close to the optimal level even when combined with the optimal prioritization policy.

□ The benefit of applying the optimal reservation policy is the most evident when a firm's capacity is at intermediate levels, as shown in Fig. 6(b). With scarce capacity, outsourcing of emergency orders and backlogging of both classes of regular orders are almost surely to happen regardless of how reservation is made. With ample capacity, most of emergency orders and regular orders will be processed with or without reservations. In these two scenarios, the expected profit resulting from *PH1st-AvgR* and *PH1st-MinR* are both close to the optimal level. When capacity is at intermediate levels, the reservation quantity affects how many orders from each class are fulfilled and hence an optimal reservation decision becomes critical.

□ The performance of *MinR* does not dominate that of *AvgR*. *AvgR* has a static reservation level while *MinR* adjusts the reservation level based on how the queue length of the PH orders compares to capacity each period. Numerical Study 2 indicates that setting a fixed reservation level is not necessarily a worse policy than adjusting the reservation level periodically in a simple way. Under the optimal policy, the reservation level can be the minimal reservation quantities: \bar{C}_{lt} and \bar{C}_{ht} , or larger quantities, and choosing which quantity to use involves examinations of several factors. *MinR* sticks to the two minimal reservation quantities and chooses between them through a simple comparison. While *MinR* is much easier to implement, it adopts a reservation quantity smaller than the optimal one from time to time. Taking the average of \bar{C}_{lt} and \bar{C}_{ht} as the reservation quantity, *AvgR* never makes the optimal reservation, but it has a reservation quantity closer to the optimal level than *MinR* does when *MinR* chooses $\bar{C}_{ht} < \bar{C}_{lt}$ and the optimal policy does not.

Overall, we find a great prioritization policy that is simple to use: *PH1st*. But, a firm will not benefit much from *PH1st* if it is not accompanied with a good reservation policy. When a firm has extreme capacity levels, *AvgR* or *MinR* can be considered as substitutes of the optimal reservation policy. When a firm has intermediate capacity levels, applying the complicated optimal policy is worthwhile.

7. Conclusion

We formulate a finite-horizon, discrete-time, dynamic programming model in which a firm decides how to make reservation for emergency demand and how to prioritize two classes of regular demand. The reservation decision is difficult because it must be made before emergency demand arrives. The prioritization decision is difficult because (1) backlogging cost accumulates over multiple periods before a unit of regular demand is satisfied while revenue is only collected upon the fulfillment of regular demand and (2) a unit of regular demand with a higher revenue may have a smaller backlogging cost.

We develop metrics that help prioritize demand, identify conditions under which one class of regular demand is dominant, and fully characterize the optimal policy. Our research suggests that a one-size-fits-all capacity allocation rule does not exist because the firm’s optimal prioritization and reservation policy is period-dependent. We also find that the existence of the dominant demand type is not automatic. When one of the regular demand classes is dominant, it is processed before the other regular demand class gets any capacity allocation. And, more dominant demand does not necessarily result in more backlogs because the firm can cut down on capacity reservation, if necessary or possible. When none of the regular demand classes is dominant, the class of regular demand with a smaller maximal loss could be processed before the other class gets any capacity allocation.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at [10.1016/j.ejor.2015.06.020](https://doi.org/10.1016/j.ejor.2015.06.020)

Appendix A. Supplementary Technical Statements

A.1. Properties of the first-order-derivative related functions

Lemma A.1. $v_{it}(z_{it}, z_{jt}, X_t)$ is continuous in z_{it} , z_{jt} and X_t , with $\frac{\partial v_{it}}{\partial z_{it}} \leq \frac{\partial v_{it}}{\partial z_{jt}} \leq -\frac{\partial v_{it}}{\partial X_t} \leq 0$. $C \geq \bar{C}_{it} \Leftrightarrow v_{it}(z_{it}, z_{jt}, z_{it} + z_{jt}) \leq 0$ and $C \neq \bar{C}_{it} \Rightarrow v_{it}(z_{it}, z_{jt}, z_{it} + z_{jt}) \neq 0$.

Lemma A.2. Given $C > \bar{C}_{it}$, $v_{it}(\bar{z}_{it}, z_{jt}, X_t) \leq 0$. More specifically, if $z_{jt}^p < z_{jt} \leq X_t$ then $v_{it}(\bar{z}_{it}, z_{jt}, X_t) < 0$; if $z_{jt}^c < z_{jt} \leq z_{jt}^p$ then $v_{it}(\bar{z}_{it}, z_{jt}, X_t) = 0$; otherwise, $m_{it} \leq c_e \Rightarrow v_{it}(\bar{z}_{it}, z_{jt}, X_t) = 0$ and $m_{it} > c_e \Rightarrow v_{it}(\bar{z}_{it}, z_{jt}, X_t) < 0$.

Lemma A.3. $\bar{v}_{it}(z_{it}, z_{jt})$ decreases in z_{it} but increases in z_{jt} . Given $z_{jt} \leq z_{jt}^p(X_t)$ and $z_{it} = \bar{z}_{it}(z_{jt}, X_t)$, if $\bar{v}_{it}(z_{it}, z_{jt}) \leq 0$ then $\bar{z}_{jt}(z_{it}, X_t) \geq z_{jt}$; and if $\bar{v}_{it}(z_{it}, z_{jt}) \geq 0$, then $\bar{z}_{jt}(z_{it}, X_t) \leq z_{jt}$.

A.2. Induction properties and their impact on dynamic losses

Since we solve the (MDP) model by backward induction, the same group of induction properties should hold for all the periods $t = 1, 2, \dots, T$.

\mathcal{IP}_t^M . **The monotonic properties of the first derivatives of $G_t(\mathbf{x}_t)$.** This set of properties aims to establish the joint concavity and sub-modularity of $G_t(\mathbf{x}_t)$ in \mathbf{x}_t . Due to the possibility of $m_{it} > c_e$ (which forces \bar{C}_{it} to 0), even with the continuous differentiable assumption on the distribution functions, $(G_t)'_i(\mathbf{x}_t)$ might not be continuous at

$x_{it} = C$ though it is continuous in x_{jt} . Hence, it is not necessarily true that $(G_t)''_{ji}(\mathbf{x}_t) = (G_t)''_{ji}(\mathbf{x}_t)$. \mathcal{IP}_t^M includes the following parts:

1. $(G_t)'_i(\mathbf{x}_t)$ decreases in both x_{it} and x_{jt} .
2. $(G_t)'_i(\mathbf{x}_t)$ decreases faster in x_{it} than in x_{jt} .
3. $(G_t)'_i(\mathbf{x}_t)$ decreases faster in x_{it} than $(G_t)'_j(\mathbf{x}_t)$ does.

\mathcal{IP}_t^B . **The boundary properties of the first derivatives of $G_t(\mathbf{x}_t)$.** This set of properties includes two elements. The first element shows how an additional unit of class i demand affects the profit. The second element compares the respective marginal impact that an additional unit has on the two classes of regular demand when their period type and final type are consistent. Specifically,

1. If $C < \bar{C}_{it}$ then $(G_t)'_i(\mathbf{x}_t) = -n_{it}$, otherwise $-n_{it} \leq (G_t)'_i(\mathbf{x}_t) \leq r_i - c_e \bar{F}_e(C)$
2. If $m_i \geq m_j$ and $m_{it} \geq m_{jt}$, then $(G_t)'_i(\mathbf{x}_t) - (G_t)'_j(\mathbf{x}_t) \leq (n_{jt} - n_{it}) \vee (r_i - r_j)$

\mathcal{IP}_t^C . **The convergency properties of the first derivatives of $G_t(\mathbf{x}_t)$.** There exists a switching curve $\phi_{it}(x_{jt})$ satisfying

$$x_{it} \geq \phi_{it}(x_{jt}) \Leftrightarrow (G_t)'_i(\mathbf{x}_t) = -n_{it} \text{ and } x_{it} \geq \phi_{it}(x_{jt}) \Rightarrow (G_t)''_{ji}(\mathbf{x}_t) = 0.$$

It has the following properties

1. $\phi_{it} \geq \phi_{i,t-1} \geq 0$ and if $C \leq \bar{C}_{it}$, $\phi_{it} = 0$
2. $\frac{\partial \phi_{it}}{\partial x_{jt}} \in [-1, 0]$, and there exists a x_{jt}^0 such that $x_{jt} \geq x_{jt}^0 \Rightarrow \frac{\partial \phi_{it}}{\partial x_{jt}} = 0$. That is, by letting $\phi_{it}(0) = \bar{\phi}_{it}$ and $\phi_{it}(x_{jt}^0) = \underline{\phi}_{it}$, $\phi_{it}(x_{jt}) \in [\bar{\phi}_{it}, \underline{\phi}_{it}]$ and $x_{jt} \geq x_{jt}^0 \Rightarrow \phi_{it}(x_{jt}) = \underline{\phi}_{it}$.

The basic idea of \mathcal{IP}_t^C is that, for any given x_{jt} , when x_{it} reaches a certain value $\phi_{it}(x_{jt})$, the two first derivatives of $G_t(\mathbf{x}_t)$ converge to two constant numbers, respectively.

Based on the above induction properties, *ML* (maximal loss) and *LDL* (limiting dynamic loss) can also be formulated as

$$m_{it} = \bar{m}_{it}(\bar{\phi}_{i,t-1}, z_{jt}) \text{ and } \bar{m}_{it}^L = \bar{m}_{it}(0, \bar{\phi}_{i,t-1}),$$

*A.3. Validation of \bar{z} , z^c , and z^**

Proposition A.4 is built on **Lemma A.1**, which examines the properties of $v_{it}(z_{it}, z_{jt}, X_t)$. The proposition validates the definition of \bar{z}_{it} in (1), and characterizes the shape of \bar{z}_{it} .

Proposition A.4. $\bar{z}_{it}(z_{jt}, X_t)$, as defined in (1), satisfies $\bar{z}_{it}(z_{jt}, X_t) = \arg \max_{(X_t - z_{jt})^+ \leq z_{it} \leq X_t - z_{jt}} V_t(\mathbf{z}_t, X_t)$. It has the following properties:

1. $\bar{z}_{it}(z_{jt}, X_t)$ is continuous in z_{jt} and X_t , with $\frac{\partial \bar{z}_{it}}{\partial z_{jt}} \in [-1, 0]$, $\frac{\partial \bar{z}_{it}}{\partial X_t} \in [0, 1]$, and $\frac{\partial \bar{z}_{it}}{\partial z_{jt}} + \frac{\partial \bar{z}_{it}}{\partial X_t} \in [-1, 0]$.
2. If $C \leq \bar{C}_{it}$, then $\bar{z}_{it}(z_{jt}, X_t) = X_t - z_{jt}$. Otherwise, two switching curves $0 \leq z_{jt}^c(X_t) \leq z_{jt}^p(X_t) \leq X_t$ can be identified, such that $0 \leq z_{jt} \leq z_{jt}^c \Rightarrow \bar{z}_{it} = X_t - (C - \bar{C}_{it}) - z_{jt}$, $z_{jt}^c < z_{jt} < z_{jt}^p \Rightarrow \bar{z}_{it} > (X_t - C + \bar{C}_{it} - z_{jt})^+$, and $z_{jt}^p \leq z_{jt} \leq X_t \Rightarrow \bar{z}_{it} = 0$. More specifically, (i) there exist two thresholds: $\bar{X}_{it}^p \geq X_{it}^p > 0$ such that,

$$0 \leq X_t \leq X_{it}^p \Rightarrow z_{jt}^p = 0, \quad X_{it}^p \leq X_t \leq \bar{X}_{it}^p \Rightarrow \frac{\partial z_{jt}^p}{\partial X_t} \in (0, 1),$$

$$X_t \geq \bar{X}_{it}^p \Rightarrow \frac{\partial z_{jt}^p}{\partial X_t} = 1.$$

- (ii) there also exist another two thresholds: $\bar{X}_{it}^c \geq X_{it}^c \geq C - \bar{C}_{it}$ such that,

$$0 \leq X_t < X_{it}^c \Rightarrow z_{jt}^c = 0, \quad X_{it}^c \leq X_t \leq \bar{X}_{it}^c \Rightarrow \frac{\partial z_{jt}^c}{\partial X_t} > 1,$$

$$X_t \geq \bar{X}_{it}^c \Rightarrow \frac{\partial z_{jt}^c}{\partial X_t} = 1.$$

$$(iii) X_{it}^c \geq C - \bar{C}_{it} \geq X_{it}^p, \bar{X}_{it}^c \geq \bar{X}_{it}^p \text{ and } z_{jt}^p(\bar{X}_{it}^p) \geq z_{jt}^c(\bar{X}_{it}^c). c_e \leq \bar{m}_{it}^L \Leftrightarrow X_{it}^c = \bar{X}_{it}^p \text{ and } z_{jt}^p(\bar{X}_{it}^p) = z_{jt}^c(\bar{X}_{it}^c).$$

The equilibrium optimal $z^e(X_t)$ defined in (2) is essentially the amount of regular demand the firm should backlog in period t , whenever possible. Proposition A.5 validates the definition of $z_t^e(X_t)$.

Proposition A.5. For any given X_t , there exists a constant δ such that $z_{1t} + z_{2t} = \delta$ for every $z_t \in Z_t^e(X_t)$, and $z_t^e(X_t)$ satisfies $z_t^e(X_t) = \arg \max_{z_t \in Z_t^e(X_t)} V_t(z_t, X_t)$. More specifically, 1. if $C \leq (\bar{C}_{1t} \wedge \bar{C}_{2t})$, $\{z_t^e(X_t)\} = \{z_t \geq 0 | z_{1t} + z_{2t} = X_t\}$; 2. if $(\bar{C}_{1t} \wedge \bar{C}_{2t}) < C \leq (\bar{C}_{1t} \vee \bar{C}_{2t})$, $m_{it} \geq m_{jt} \Leftrightarrow z_{it}^e = 0$ and $z_{jt}^e = X_t$; 3. if $C > (\bar{C}_{1t} \vee \bar{C}_{2t})$, $X_t - C + (\bar{C}_{1t} \vee \bar{C}_{2t}) \leq z_{1t}^e + z_{2t}^e \leq X_t$.

After obtaining $\bar{z}_{it}(z_{jt}, X_t)$ and subsequently $z_t^e(X_t)$, we can apply (3) to identify the optimal amount of regular demand to be backlogged, namely z_t^* . Proposition A.6 justifies this procedure.

Proposition A.6. $z_t^*(x_t)$, as defined in (3), satisfies $z_t^*(x_t) = \arg \max_{z_t \in Z_t(x_t)} V_t(z_t, X_t)$, with $\frac{\partial z_t^*}{\partial x_t} \in [0, 1]$, and $\left| \frac{\partial z_t^*}{\partial x_t} \right| \leq 1$.

The establishment of Proposition A.6 allows us to explore the properties of the optimal profit function $G_t(x_t)$ in Proposition A.7, which links $G_t(x_t)$ to the induction properties.

Proposition A.7. Given $z_t^*(x_t)$ established in (3), all the three induction properties also hold for period t . That is, $G_t(x_t)$ has the properties \mathcal{IP}_t^M , \mathcal{IP}_t^B , and \mathcal{IP}_t^C .

A.4. Structural results for the Only-High and Only-Low regions

For $\mathcal{I} = \mathcal{F}, \mathcal{M}$, denote by $\mathcal{H}_t^{\mathcal{I}}(X_t)$ a point on $\mathcal{H}_t^{\mathcal{I}}$ with coordinates $\mathcal{H}_{it}^{\mathcal{I}}(X_t)$ and $\mathcal{H}_{jt}^{\mathcal{I}}(X_t)$, and denote by $\mathcal{L}_t^{\mathcal{I}}(X_t)$ a point on $\mathcal{L}_t^{\mathcal{I}}$ with coordinates $\mathcal{L}_{it}^{\mathcal{I}}(X_t)$ and $\mathcal{L}_{jt}^{\mathcal{I}}(X_t)$. Note that $\mathcal{H}_{it}^{\mathcal{I}}(X_t) + \mathcal{H}_{jt}^{\mathcal{I}}(X_t) = X_t$ and $\mathcal{L}_{it}^{\mathcal{I}}(X_t) + \mathcal{L}_{jt}^{\mathcal{I}}(X_t) = X_t$.

Lemma A.8. Suppose $C > (\bar{C}_{it} \vee \bar{C}_{jt})$ and $m_{it} \geq m_{jt}$. Either (i) $m_i \geq m_j$ or (ii) $m_i < m_j$ and $\bar{m}_{it}^L \geq m_{jt}$. We have

$$\mathcal{H}_{it}^{\mathcal{F}}(X_t) = X_t - z_{jt}^p(X_t), \mathcal{H}_{jt}^{\mathcal{F}}(X_t) = z_{jt}^p(X_t) \text{ and } \mathcal{H}_{it}^{\mathcal{M}}(X_t) = X_t - z_{jt}^c(X_t), \mathcal{H}_{jt}^{\mathcal{M}}(X_t) = z_{jt}^c(X_t).$$

They have the following properties:

- $X_t \leq X_{it}^p \Rightarrow \mathcal{H}_{it}^{\mathcal{F}} = X_t, X_{it}^p < X_t < \bar{X}_{it}^p \Rightarrow \frac{\partial \mathcal{H}_{it}^{\mathcal{F}}}{\partial X_t} \in (0, 1)$, and $\bar{X}_{it}^p \leq X_t \Rightarrow \frac{\partial \mathcal{H}_{it}^{\mathcal{F}}}{\partial X_t} = 0, \mathcal{H}_{it}^{\mathcal{F}} \geq C - \bar{C}_{it}$.
- $X_t \leq X_{it}^c \Rightarrow \mathcal{H}_{it}^{\mathcal{M}} = X_t, X_{it}^c < X_t < \bar{X}_{it}^c \Rightarrow \frac{\partial \mathcal{H}_{it}^{\mathcal{M}}}{\partial X_t} < 0$, and $\bar{X}_{it}^c \leq X_t \Rightarrow \frac{\partial \mathcal{H}_{it}^{\mathcal{M}}}{\partial X_t} = 0$.
- If $c_e \geq \bar{m}_{it}^L$, then $\mathcal{H}_{jt}(X_t^{MR}) = \mathcal{H}_{jt}^{\mathcal{F}}(\bar{X}_{it}^p) \geq \mathcal{H}_{jt}^{\mathcal{M}}(\bar{X}_{it}^c)$; if $\bar{m}_{it}^L > c_e \geq m_{jt}$, then $\mathcal{H}_{jt}(X_t^{MR}) \geq \mathcal{H}_{jt}^{\mathcal{F}}(\bar{X}_{it}^p) = \mathcal{H}_{jt}^{\mathcal{M}}(\bar{X}_{it}^c)$; and if $c_e < m_{jt}$, then $\mathcal{H}_{jt}(X_t^{MR}) \geq \mathcal{H}_{jt}^{\mathcal{F}}(\bar{X}_{it}^p) = \mathcal{H}_{jt}^{\mathcal{M}}(\bar{X}_{it}^c)$ and $\mathcal{H}_{it}^{\mathcal{F}}(\bar{X}_{it}^p) = C$.

Lemma A.9. Suppose $C > (\bar{C}_{it} \vee \bar{C}_{jt})$. Given $m_{it} \geq m_{jt}, m_i < m_j$, and $\bar{m}_{it}^L < m_{jt}$, we have $\mathcal{H}_{it}^{\mathcal{F}} + \mathcal{H}_{jt}^{\mathcal{F}} = X_t$ and $\mathcal{H}_{it} \leq \mathcal{H}_{it}^{\mathcal{F}}$. If $\bar{m}_{it}^S \leq \bar{m}_{it}^S$, then $\mathcal{H}_{it} = \mathcal{H}_{it}^{\mathcal{F}}$; otherwise, $X_t < X_{it}^{PS} \Rightarrow \mathcal{H}_{it}^{\mathcal{F}} = X_t - z_{jt}^p(X_t)$ and $X_t \geq X_{it}^{PS} \Rightarrow \mathcal{H}_{it}^{\mathcal{F}} = \mathcal{H}_{it}$, so that 1. $X_{it}^{FS} < X_{it}^p$ and $X_t \leq X_{it}^p \Rightarrow \mathcal{H}_{it}^{\mathcal{F}} = X_t$; 2. if $c_e \geq \bar{m}_{it}^E$, then $X_{it}^p < X_t < X_{it}^{PS} \Rightarrow \frac{\partial \mathcal{H}_{it}^{\mathcal{F}}}{\partial X_t} \in (0, 1)$; 3. if $c_e < \bar{m}_{it}^E$, then $\bar{X}_{it}^p \leq X_{it}^{MR} \leq X_{it}^{PS}, X_{it}^p < X_t < \bar{X}_{it}^p \Rightarrow \frac{\partial \mathcal{H}_{it}^{\mathcal{F}}}{\partial X_t} \in (0, 1)$, and $\bar{X}_{it}^p \leq X_t < X_{it}^{PS} \Rightarrow \mathcal{H}_{it}^{\mathcal{F}} = C$.

Lemma A.10. Suppose $C > (\bar{C}_{it} \vee \bar{C}_{jt})$. Given $m_{it} \geq m_{jt}, m_i < m_j$, and $\bar{m}_{it}^L < m_{jt}$, we have $\mathcal{H}_{it}^{\mathcal{M}} + \mathcal{H}_{jt}^{\mathcal{M}} = X_t$ and $\mathcal{H}_{it}^{\mathcal{F}} \leq \mathcal{H}_{it}^{\mathcal{M}}$. More specifically,

- if $c_e \geq m_{jt}$, then $\mathcal{H}_{it}^{\mathcal{M}} = X_t - z_{jt}^c(X_t)$, so that (i) $X_t \leq X_{it}^c \Rightarrow \mathcal{H}_{it}^{\mathcal{M}} = X_t$; (ii) $X_{it}^c < X_t < \bar{X}_{it}^c \Rightarrow \frac{\partial \mathcal{H}_{it}^{\mathcal{M}}}{\partial X_t} < 0$; (iii) $\bar{X}_{it}^c \leq X_t \Rightarrow \frac{\partial \mathcal{H}_{it}^{\mathcal{M}}}{\partial X_t} = 0$.
- if $c_e < m_{jt}$, then $X_t < X_{it}^{MR} \Rightarrow \mathcal{H}_{it}^{\mathcal{M}} = X_t - z_{jt}^c(X_t)$ and $X_t \geq X_{it}^{MR} \Rightarrow \mathcal{H}_{it}^{\mathcal{M}} = \mathcal{H}_{it}^{\mathcal{F}}$. Given $X_t < X_{it}^{MR}$, (i) if $\bar{m}_{it}^S \leq \bar{m}_{it}^S$ and $c_e \geq \bar{m}_{it}^E$, then $X_t \leq X_{it}^c \Rightarrow \mathcal{H}_{it}^{\mathcal{M}} = X_t, X_t > X_{it}^c \Rightarrow \frac{\partial \mathcal{H}_{it}^{\mathcal{M}}}{\partial X_t} < 0$; (ii) if $\bar{m}_{it}^S \leq \bar{m}_{it}^S$ and $c_e < \bar{m}_{it}^E$, then $\bar{X}_{it}^c < X_{it}^{MR}, X_t \leq X_{it}^c \Rightarrow \mathcal{H}_{it}^{\mathcal{M}} = X_t, X_{it}^c < X_t < \bar{X}_{it}^c \Rightarrow \frac{\partial \mathcal{H}_{it}^{\mathcal{M}}}{\partial X_t} < 0$, and $\bar{X}_{it}^c < X \Rightarrow \mathcal{H}_{it}^{\mathcal{M}} = C$; (iii) If $\bar{m}_{it}^S > \bar{m}_{it}^S$ and $c_e < \bar{m}_{it}^E$, then $\mathcal{H}_{it}^{\mathcal{M}} = \mathcal{H}_{it}^{\mathcal{F}}$.

Lemma A.11. Suppose $C > (\bar{C}_{it} \vee \bar{C}_{jt})$. Given $m_{it} \geq m_{jt}, m_i < m_j$, and $m_{jt} > \bar{m}_{it}^L$, we have $\mathcal{L}_{it}^{\mathcal{F}} + \mathcal{L}_{jt}^{\mathcal{F}} = X_t$ and $\mathcal{L}_{jt} \leq \mathcal{L}_{jt}^{\mathcal{F}}$. If $\bar{m}_{it}^S \leq \bar{m}_{it}^S$, then $\mathcal{L}_{it} = \mathcal{L}_{it}^{\mathcal{F}}$; otherwise, $X_t < X_{it}^{PS} \Rightarrow \mathcal{L}_{it}^{\mathcal{F}} = X_t - z_{jt}^p(X_t)$ and $X_t \geq X_{it}^{PS} \Rightarrow \mathcal{L}_{it} = \mathcal{L}_{it}^{\mathcal{F}}$, so that (1) $X_{it}^{FS} < X_{it}^p$ and $X_t \leq X_{it}^p \Rightarrow \mathcal{L}_{it}^{\mathcal{F}} = X_t$; (2) if $c_e \geq \bar{m}_{it}^E$, then $X_{it}^p < X_t < X_{it}^{PS} \Rightarrow \frac{\partial \mathcal{L}_{it}^{\mathcal{F}}}{\partial X_t} \in (0, 1)$; (3) if $c_e < \bar{m}_{it}^E$, then $\bar{X}_{it}^p \leq X_{it}^{MR} \leq X_{it}^{PS}, X_{it}^p < X_t < \bar{X}_{it}^p \Rightarrow \frac{\partial \mathcal{L}_{it}^{\mathcal{F}}}{\partial X_t} \in (0, 1)$, and $\bar{X}_{it}^p \leq X_t < X_{it}^{PS} \Rightarrow \mathcal{L}_{it}^{\mathcal{F}} = C$.

Lemma A.12. Suppose $C > (\bar{C}_{it} \vee \bar{C}_{jt})$. Given $m_{it} \geq m_{jt}, m_i < m_j$, and $m_{jt} > \bar{m}_{it}^L$, we have

$$X_t < X_{it}^{MR} \Rightarrow \mathcal{L}_{it}^{\mathcal{M}} = X_t - z_{jt}^c(X_t) \text{ and } X_t \geq X_{it}^{MR} \Rightarrow \mathcal{L}_{it}^{\mathcal{M}} = \mathcal{L}_{it}^{\mathcal{F}},$$

satisfying $\mathcal{L}_{it}^{\mathcal{M}} + \mathcal{L}_{jt}^{\mathcal{M}} = X_t$ and $\mathcal{L}_{jt}^{\mathcal{F}} \leq \mathcal{L}_{jt}^{\mathcal{M}}$. More specifically, for $X_t < X_{it}^{MR}$, we have (1) given $\bar{m}_{it}^S \leq \bar{m}_{it}^S$, if $c_e \geq \bar{m}_{it}^E$, then $X_t \leq X_{it}^c \Rightarrow \mathcal{L}_{it}^{\mathcal{M}} = X_t$, and $X_t > X_{it}^c \Rightarrow \frac{\partial \mathcal{L}_{it}^{\mathcal{M}}}{\partial X_t} < 0$; (2) given $\bar{m}_{it}^S \leq \bar{m}_{it}^S$, if $c_e < \bar{m}_{it}^E$, then $\bar{X}_{it}^c < X_{it}^{MR}, X_t \leq X_{it}^c \Rightarrow \mathcal{L}_{it}^{\mathcal{M}} = X_t, X_{it}^c < X_t < \bar{X}_{it}^c \Rightarrow \frac{\partial \mathcal{L}_{it}^{\mathcal{M}}}{\partial X_t} < 0$, and $\bar{X}_{it}^c < X \Rightarrow \mathcal{L}_{it}^{\mathcal{M}} = C$; (3) given $\bar{m}_{it}^S > \bar{m}_{it}^S$, if $c_e < \bar{m}_{it}^E$, then $\mathcal{L}_{it}^{\mathcal{M}} = \mathcal{L}_{it}^{\mathcal{F}}$.

References

Alexopoulos, K., Mourtzis, D., Papakostas, N., & Chrysolouris, G. (2007). Desyma – Assessing flexibility for the lifecycle of manufacturing systems. *International Journal of Production Research*, 45, 1683–1694.

Alexopoulos, K., Papakostas, N., Mourtzis, D., & Chrysolouris, G. (2011). A method for comparing flexibility performance for the lifecycle of manufacturing systems under capacity planning constraints. *International Journal of Production Research*, 49, 3307–3317.

Carr, S., & Duenyas, I. (2000). Optimal admission control and sequencing in a make-to-stock/make-to-order production system. *Operations Research*, 48, 709–720.

Chen, R., & Robinson, L. W. (2010). A comparison of traditional and open-access policies for appointment scheduling. *Manufacturing and Service Operations Management*, 12, 330–346.

deKoster, R., Le-Duc, T., & Roodbergen, K. J. (2007). Design and control of warehouse order picking: A literature review. *European Journal of Operational Research*, 182, 481–501.

Dobson, G., Hasija, S., & Pinker, E. J. (2011). Reserving capacity for urgent patients in primary care. *Production and Operations Management*, 20, 456–473.

Frank, K. C., Zhang, R. Q., & Duenyas, I. (2003). Optimal policies for inventory systems with priority demand classes. *Operations Research*, 51, 993–1002.

Gao, L., Xu, S. H., & Ball, M. O. (2012). Managing an available-to-promise assembly system with dynamic short-term pseudo-order forecast. *Management Science*, 58, 770–790.

Gayon, J. P., de Vericourt, F., & Karaesmen, F. (2009). Stock rationing in an m/e/1 multi-class make-to-stock queue with backorders. *IIE Transactions*, 41, 1096–1109.

Geng, N., & Xie, X. (2012). Optimizing contracted resource capacity with two advance cancellation modes. *European Journal of Operational Research*, 221, 501–512.

Gerchak, Y., Gupta, D., & Henig, M. (1996). Reservation planning for elective surgery under uncertain demand for emergency surgery. *Management Science*, 42, 321–334.

Green, L. V. (2012). The vital role of operations analysis in improving healthcare delivery. *Manufacturing and Service Operations Management*, 14, 488–494.

Green, L. V., Savin, S., & Wang, B. (2006). Managing patient service in a diagnostic medical facility. *Operations Research*, 54, 11–25.

Gupta, D., & Wang, L. (2007). Capacity management for contract manufacturing. *Operations Research*, 55, 367–377.

Gupta, D., & Wang, L. (2008). Revenue management for a primary-care clinic in the presence of patient choice. *Operations Research*, 56, 576–592.

Ha, A. Y. (1997). Stock-rationing policy for a make-to-stock production system with two priority classes and backordering. *Naval Research Logistics*, 44, 458–472.

- Huang, B., & Iravani, S. M. R. (2008). A make-to-stock system with multiple customer classes and batch ordering. *Operations Research*, 56, 1312–1320.
- Iravani, S. M. R., Liu, T., & Simchi-Levi, D. (2012). Optimal production and admission policies in make-to-stock/make-to-order manufacturing systems. *Production and Operations Management*, 21, 224–235.
- Makris, S., & Chrysolouris, G. (2010). Customer's behavior modeling for manufacturing planning. *International Journal of Computer Integrated Manufacturing*, 23, 619–629.
- May, J. H., Spangler, W. E., Strum, D. P., & Vargas, L. G. (2011). The surgical scheduling problem: Current research and future opportunities. *Production and Operations Management*, 20, 392–405.
- McGill, J. I., & van Ryzin, G. J. (1999). Revenue management: Research overview and prospects. *Transportation Science*, 33, 233–256.
- Min, D., & Yih, Y. (2010). Scheduling elective surgery under uncertainty and downstream capacity constraints. *European Journal of Operational Research*, 206, 642–652.
- Mourtzis, D., Doukas, M., & Psarommatis, F. (2013). Manufacturing network design for mass customization using a genetic algorithm and an intelligent search method. *Procedia CIRP*, 7, 37–42.
- Mourtzis, D., Doukas, M., Psarommatis, F., Giannoulis, C., & Michalos, G. (2014). A web-based platform for mass customization and personalization. *CIRP Journal of Manufacturing Science and Technology*, 7, 112–128.
- Papakostas, N., & Mourtzis, D. (2007). An approach for adaptability modeling in manufacturing – Analysis using chaotic dynamics. *Annals of CIRP – Manufacturing Technology*, 56, 491–494.
- Patrick, J., & Puterman, M. L. (2007). Improving resource utilization for diagnostic services through flexible inpatient scheduling: A method for improving resource utilization. *Journal of the Operational Research Society*, 58, 235–245.
- Patrick, J., Puterman, M., & Queyranne, M. (2008). Dynamic multi-priority patient scheduling for a diagnostic resource. *Operations Research*, 56, 1507–1525.
- Rahman, A. A. A., & Seliger, G. (2013). Integrated control system simulation for supporting changes of routing strategy in an automated material flow system. *Procedia CIRP*, 7, 407–412.
- Ramirez-Nafarrate, A., Hafizoglu, A. B., Gel, E. S., & Fowler, J. W. (2014). Optimal control policies for ambulance diversion. *European Journal of Operational Research*, 236, 298–312.
- Talluri, K. T., & van Ryzin, G. J. (2004). *The theory and practice of revenue management*. Boston: Kluwer Academic Publishers.
- de Vericourt, F., Karaesmen, F., & Dallery, Y. (2002). Optimal stock allocation for a capacitated supply system. *Management Science*, 48, 1486–1501.
- Wang, H., Liang, X., Sethi, S., & Yan, H. (2014). Inventory commitment and prioritized backlogging clearance with alternative delivery lead times. *Production and Operations Management*, 23, 1227–1242.