Responsible Sourcing and Supply Chain Risk Management

by

Lu Huang

Business Administration
Duke University

Date: __________________________

Approved:

___________________________
Jing-Sheng Song, Supervisor

___________________________
Robert Swinney

___________________________
Fernando Bernstein

___________________________
Peng Sun

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Business Administration
in the Graduate School of Duke University
2015
ABSTRACT

Responsible Sourcing and Supply Chain Risk Management

by

Lu Huang

Business Administration
Duke University

Date: __________________________

Approved:

______________________________
Jing-Sheng Song, Supervisor

______________________________
Robert Swinney

______________________________
Fernando Bernstein

______________________________
Peng Sun

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Business Administration
in the Graduate School of Duke University
2015
Abstract

With the extensive use of outsourcing and more frequent technological innovations, global supply chains become vastly stretched and dynamic networks. As a result, firms face significant challenges in managing their increasingly fragile supply networks and responding to rapidly changing demand evolutions. In this dissertation, we analyze three aspects of these challenges and summarize the findings in three essays. The first essay considers a firm’s problem of managing social and environmental non-compliance risks at its sub-tier suppliers. We identify the conditions under which the firm should delegate the management of sub-tier suppliers’ non-compliance to its direct suppliers, instead of direct control. In the second essay, we analyze the firm’s strategy to deal with random demand surges. We develop a new demand model that captures important non-Markovian characteristics of possible demand surge trajectories and derive the optimal joint safety stock and reactive capacity strategy. We further study the impact of surge demand characteristics (e.g., frequency, intensity, duration, and shape). In the third essay, we examine a dynamic customer-base management problem for a firm with finite capacity, when its customers are prone to disruption and retention risks. We show that the optimal acquisition policy is an add-up-to policy and derive the firm’s optimal capacity allocation policy when capacity shortage occurs. Collectively, the studies in this dissertation provide useful modeling ideas, decision tools, insights, and guidance for firms to build resilient supply chains from both the supply and demand sides.
Contents

Abstract iv
List of Tables viii
List of Figures ix
List of Abbreviations and Symbols x
Acknowledgements xiii

1 Introduction 1

2 Managing Social Responsibility in Multi-tier Supply Chains 6
   2.1 Introduction .................................................. 6
   2.2 Literature Review ........................................... 11
   2.3 Model .......................................................... 14
   2.4 The Manufacturer’s Optimal Strategy .................... 20
   2.5 External Stakeholder Pressure .............................. 27
   2.6 Penalty Contracts ............................................. 32
   2.7 Complementary Effort ........................................ 37
   2.8 Conclusion .................................................... 39

3 Supply Chain Planning for Random Demand Surges: Reactive Capacity and Safety Stock 43
   3.1 Introduction .................................................... 43
   3.2 Literature Review ............................................. 48
3.3 Model ................................................................. 50
  3.3.1 Demand Process ........................................... 50
  3.3.2 Problem Formulation ..................................... 53
3.4 Optimization .................................................. 56
  3.4.1 Optimal Deployment Policy ............................... 56
  3.4.2 Problem Transformation ................................ 59
  3.4.3 Optimal Planning Policy ................................. 61
3.5 Impact of Product and Demand Characteristics ............. 65
  3.5.1 Product Characteristics .................................. 65
  3.5.2 Demand Characteristics .................................. 67
  3.5.3 Sensitivity Results Summary ............................. 76
3.6 Concluding Remarks .......................................... 76

4 Dynamic Customer Base Management: Acquisition, Retention, and Risk Diversification 79
  4.1 Introduction .................................................. 79
  4.2 Literature Review ............................................ 83
  4.3 Model .......................................................... 86
    4.3.1 Supply Chain Structure ................................. 86
    4.3.2 Customer Risk Profile .................................. 87
    4.3.3 Supplier’s Problem ...................................... 89
  4.4 Infinite Horizon Problem .................................... 91
    4.4.1 Dynamic Programming Formulation ..................... 91
    4.4.2 Optimal Strategy Analysis ............................... 93
    4.4.3 Sensitivity Analysis ..................................... 98
  4.5 Finite Horizon Problem ...................................... 102
    4.5.1 Dynamic Programming Formulation ..................... 102
4.5.2 Optimal Strategy for One Type of Customers ($k = 1$) ... 103
4.5.3 Optimal Strategy for Two Types of Customers ($k = 2$) ... 104
4.6 Insights and Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . 105

A Appendix 108

Bibliography 152

Biography 158
List of Tables

2.1  The impact of external stakeholder pressure on equilibrium responsibility effort under an exponential risk function $\alpha(x) = \beta e^{-x}$. . . . . .  31

2.2  Impact of model parameters on the coordinating contract $(w^*, \varphi^*)$ . .  34

3.1  Summary of product and surge demand characteristics impact on the optimal planning policy. * indicates maintaining Lorenz order equality.  77
# List of Figures

2.1 Supply chain overview ................................................. 18  
2.2 Sequence of events ...................................................... 19  
2.3 The Manufacturer’s optimal wholesale price and strategy ............. 26  
2.4 Greater external stakeholder pressure leading to lower responsibility effort. ........................................................... 30  
3.1 Depiction of possible surge demand paths ............................. 51  
3.2 Example of optimal deployment policy. Here, the dotted line triangle denotes the building of inventory from reactive capacity that is subsequently used after the safety stock depletes. ............................. 57  
3.3 Depiction of the critical time determination for the increasing and stationary case (left) and the uni-modal or decreasing case (right). ............................. 60  
3.4 Optimal policy region under case $c < w$ ............................. 63  
3.5 The impact of the shortage penalty on the optimal policy ............. 67  
3.6 Impact of demand duration increases under dilation order (left) and under stochastic order with no control of uncertainty (right) ............. 70  
4.1 Supply chain structure .................................................... 86  
4.2 Timing of event ............................................................ 92  
4.3 The impact of disruption risk $p$ and demand uncertainty $D$ ........... 101
List of Abbreviations and Symbols

Symbols for Chapter 2

As in convention, the superscript * is used to represent the corresponding optimal quantity.

- $e_i$ non-compliance effort from Tier $i$ company
- $k_i$ unit effort cost at Tier $i$ company
- $p$ unit sale price of Tier 0 company
- $w$ unit sale price of Tier 1 company
- $c$ unit production cost of Tier 1 company
- $D/d$ demand at Tier 0 company
- $\alpha (e)$ the probability of non-compliance at Tier 0 company
- $\gamma$ the proportion of demand loss from non-compliance
- $\varphi$ the one-time penalty at Tier 1 company from Tier 0 company
- $a_i$ indirect financial loss of Tier $i$ company from non-compliance
- $L_i (w)$ the total non-compliance loss of Tier $i$

Symbols for Chapter 3

- $\lambda$ rate of entering surge state, $\nu = \frac{1}{\lambda}$ expected duration in normal state
- $(\Omega, \mathcal{F}, \mathbb{P})$ probability space for all possible demand sample paths,
- $T (\omega) | \omega \in \Omega$ sample demand surge duration, $\tau$ expected duration in surge state
- $X (t, \omega) | \omega \in \Omega$ sample demand surge magnitude at time $t$
\{X(t, \omega), 0 \leq t \leq T(\omega)\} \quad \text{sample demand surge path}

\begin{align*}
t_p & \quad \text{peak time, } \beta = \frac{t_p(\omega)}{T(\omega)} \text{ as peak time percentage} \\
T'(\omega) & \quad \text{sample path critical time} \\
h & \quad \text{safety stock holding cost per unit/time} \\
r & \quad \text{reactive capacity reservation cost per unit/time} \\
c & \quad \text{safety stock unit purchase cost} \\
w & \quad \text{reactive capacity unit purchase cost} \\
y_m(t)/y_\mu(t) & \quad \text{deployment rate at time } t \text{ from safety stock/reactive capacity} \\
p_\mu(t) & \quad \text{production rate at time } t \text{ from reactive capacity} \\
(m, \mu) & \quad \text{safety stock and capacity reservation level} \\
\alpha & \quad \text{service level} \\
b & \quad \text{unit penalty cost for lost demand} \\
C(m, \mu) & \quad \text{optimal expected deployment cost under service level constraint} \\
C^p(m, \mu) & \quad \text{optimal expected deployment cost under unit penalty cost constraint} \\
G(m,\mu) & \quad \text{expected total cost rate under service level constraint} \\
G^p(m,\mu) & \quad \text{expected total cost rate under unit penalty cost constraint} \\
r^*(h)/h^*(r) & \quad \text{optimal policy switching curves} \\
\gamma_m & \quad \text{the proportion of expected available inventory from safety stock, used to measure the preference of safety stock}
\end{align*}

Symbols for Chapter 4

\begin{align*}
i & \in \{1, 2, \ldots, k\} \quad \text{type index} \\
j & \in \{1, 2, \ldots, n_i\} \quad \text{customer index} \\
K & \quad \text{seller's capacity limit} \\
r_i & \quad \text{unit profit margin for type } i \text{ customer} \\
\ell_i & \quad \text{unit cost to acquire a new type } i \text{ customer}
\end{align*}
β the time discount parameter

\( n^t_i \) number of type \( i \) customer, \( \mathbf{n}^t = (n^t_1, n^t_2, ..., n^t_k) \)

\( p_i \) type \( i \) customers’ disruption probability

\( a^t_{ij} \) the \( j \)-th type \( i \) customer’s availability, with \( a^t_{ij} \sim \text{Bernoulli} (1 - p_i) \)

\( \mathbf{a}^t_i = (a^t_{ij})_{j=1}^{n^t_i} \) one-period availability vector of type \( i \) customers, \( \mathbf{a}^t = (\mathbf{a}^t_1, ..., \mathbf{a}^t_k) \)

\( d^t_{i,j} \) order quantity random variable from \( j \)-th type \( i \) customer when it is not disrupted

\( \mathbf{d}^t_i = (d^t_{i,j})_{j=1}^{n^t_i} \) one-period order vector of type \( i \) customers, \( \mathbf{D}^t = (\mathbf{D}^t_1, \mathbf{D}^t_2, ..., \mathbf{D}^t_k) \)

\( q^t_{i,j} \) allocation capacity to \( j \)-th type \( i \) customer, \( 0 \leq q^t_{i,j} \leq a^t_{ij}d^t_{ij} \)

\( f^t_{i,j} \) fill rate of \( j \)-th type \( i \) customer.

\( \mathbf{f}^t_i = (f^t_{i,j})_{j=1}^{n^t_i} \) one-period fill rate vector of type \( i \) customers, \( \mathbf{f}^t = (\mathbf{f}^t_1, \mathbf{f}^t_2, ..., \mathbf{f}^t_k) \)

\( s^t_{ij} \) the \( j \)-th type \( i \) customer’s retention, with \( s^t_{ij} \sim \text{Bernoulli} (\alpha_i (f^t_{ij}) ) \)

\( \mathbf{s}^t_i = (s^t_{ij})_{j=1}^{n^t_i} \) one-period retention vector of type \( i \) customers, \( \mathbf{s}^t = (\mathbf{s}^t_1, ..., \mathbf{s}^t_k) \)

\( \alpha_i (f^t_{ij}) \) type \( i \) customers’ retention probability
Acknowledgements

I would like to express my deepest appreciation and thanks to my advisor professor Jing-Sheng Song, for her continuous support, guidance and encouragement during my PhD study and research. She is the nicest advisor with immense knowledge, enthusiasm, patience and wisdom. Without her supervision and persistent help, this dissertation would not have been possible.

I would also like to thank my coauthors professor Jordan Tong and professor Robert Swinney, for their valuable and selfless help and input in writing the essays. Their scientific advice and knowledge and many insightful discussions is essential for the essays and I am very grateful to that. I would also like to thank professor Fernando Bernstein and professor Peng Sun for serving as my committee members, and all professors in my department for enlightening me with their work and teaching.

A special thank you to my family, words cannot express how grateful I am. Your love and support for me was what sustained me thus far. Last but not least, I would like to thank all of my friends who supported and encouraged me to strive towards my goal.
This thesis consists of three essays addressing timely issues arising in global supply chains, including responsible sourcing and supply chain business-customer base and random demand surge risk management.

The trend of globalization has significantly changed supply chain structures. With easy access to abundant physical and intellectual resources all over the world at a competitive price, companies drastically increase their procurement activities, both in scale and in scope. However, with the extensive use of outsourcing, the supply chain becomes a vastly stretched network, in which a company’s normal operation depends heavily on those of its suppliers and business customers. Any malfunction at its suppliers may severely disrupt the firm’s operation. Meanwhile, the rapidly changing market, driven by technological innovation and exposure to natural and man-made disasters, has introduced significant uncertainty from the demand side for firms to handle. Together, a great deal of new and important challenges in managing more fragile supply networks and responding to more rapidly changing global demand evolution, have emerged.

There have been many best practices in industry to deal with these challenges,
and academicians have made tangible progress in preserving these good ideas by formalizing them and generalizing them. While significant progress has been made in this endeavor in the last decade or so, there are a lot more to do, which sets the goal of our research. This dissertation focuses on designing resilient supply chains through strategic choices and operational levers to hedge against extreme supply and demand risks. More specifically, we study the following three topics, reported in three essays, corresponding to Chapters 2-4 of the dissertation.

**Essay 1 (Chapter 2): Managing Social Responsibility in Multi-Tier Supply Chains**

From the supply side, firms face challenges to enforce its social and environmental standards at its vast network of suppliers. When social and environmental non-compliance accidents are reported to the public, the focal firm suffers significantly from the brand equity, stock price drop and market loss. Although majority of companies have already set up policies to enforce social compliance standards at their direct suppliers, there are many instances showing that such policies may not be enough to mitigate the risk at upper tier suppliers. Indeed, the causes of several recent high profile public tragedies, such as the factory fire in Bangladesh, traced back to the violations at the second-tier suppliers. Moreover, studies have found that in the extended global supply chains, larger and more severe risks lie actually in the sub-tier suppliers.

Motivated by these observations, in Chapter 2 we develop a model of a three-tier supply chain in which a Tier 2 supplier sells to a Tier 1 supplier, which in turn sells to a downstream (Tier 0) firm. The Tier 2 supplier potentially violates social and environmental standards, resulting in direct demand loss and indirect financial loss for the Tier 0 firm. Both the Tier 1 supplier and the Tier 0 firm can exert effort to improve the responsibility of the Tier 1 supplier and reduce the likelihood of a responsibility violation. We show that when the efforts from Tiers 0 and 1 are substitutable, the optimal strategy for manufacturer is one of extremes: either direct
control (only Tier 0 exerts effort) or delegation (Tier 0 induces Tier 1 to exert effort via contractual parameters) is optimal, but it is never optimal for both firms to exert effort simultaneously. Interestingly, we show the direct control strategy may be optimal for the Tier 0 firm even if it is more expensive for that firm to exert effort to improve responsibility than the Tier 1 firm. Given the potentially large number of higher tier suppliers, it may be prohibitively complex for a downstream firm to employ a direct control strategy in practice, meaning that delegation may be necessary even when it is not optimal; in these cases, we investigate additional mechanisms to improve the performance of the delegation strategy. Lastly, we consider the impact of pressure from external stakeholders (consumers, NGOs, and governments) on the responsibility level of the supply chain. Surprisingly, we show that more pressure from these stakeholders may actually decrease supply chain responsibility, especially when that pressure does not directly fall on the downstream firm.

Essay 2 (Chapter 3): Supply Chain Planning for Random Demand Surges: Safety Stock and Reactive Capacity

From the demand side, with global customers and rapid product and technological innovations, organizations have increased need to cope with extreme demand risks. In Chapter 3, we address a firm’s decision facing demand surges: random yet significant increases in demand in an otherwise relatively more stable demand environment. To build supply chain capabilities to manage demand surges, firms and organizations plan by investing in two fundamentally different sourcing strategies – reactive capacity and safety stock. We develop a new demand model that modifies traditional Markov-modulated demand models to capture important non-Markov characteristics of possible random demand surge trajectories. Correspondingly, to capture this non-Markov surge demand trajectories and the immediate nature of surge demand satisfaction, our service level is defined as the probability of demand satisfaction at each time point during demand surges. We then propose a stylized
representation of reactive capacity and safety stock to capture key tradeoffs between each strategy. The objective is to minimize costs subject to a service level constraint.

Subsequently, we analytically derive the optimal planning strategy by first solving the optimal deployment policy for any plan, making a key service level constraint simplification that allows us to tractably incorporate our surge demand model. Finally, through stochastic comparisons, we show how the magnitude and predictability of surge demand characteristics (e.g., frequency, intensity, duration, and shape) and the cost profiles of each strategy impact the optimal joint planning policy. We show that longer and more uncertain demand surge duration prefers reactive capacity, while earlier surge peak and more volatile demand during duration favors safety stock. Interestingly, demand intensity favors neither safety stock nor capacity reservation. Moreover, higher service level may decrease safety stock level, but never reactive capacity level. Together, these results establish a useful framework to guide supply chain planning decisions to manage random surge demands.

Essay 3 (Chapter 4): Dynamic Customer Base Management: Acquisition, Retention, and Risk Diversification

In Chapter 4, we study a firm’s dynamic customer base management in a business to business setting. The firm has finite capacity and serves multiple customer segments, with varying volumes, profit margins, loyalty and risk profiles. In a globally changing environment, the firm faces three types of risks from its customers. The first type is the external temporary disruption risk, triggered by disaster events such as natural disasters, factory fire, labor strike, etc. Once disrupted, the customers’ operations stop and therefore the firm loses orders from these customers. The second type is customers’ retention risk if the customers are not satisfied with the firm’s service level. This risk is permanent in the sense that once the customers leave, the firm loses their business forever. The third type is the usual inherent uncertainty of customers’ orders. To hedge against disruption and demand uncertainty, the firm
can acquire more customers across different segments to secure its capacity utilization. However, this may lead to lower service level and therefore higher customers’ retention risk. To guild firms to effectively make these decisions, we explore the following questions: what should be the optimal customer base? How should the firm optimally rationalize capacity under capacity shortage?

We show that, the optimal acquisition policy is an add-up-to policy for each type of customers. The up-to levels are increasing in the customers’ demand uncertainty, implying the firm should focus more on acquisition strategy to diversify the demand risks when the demand are more uncertain. However, the up-to levels are not necessarily increasing in the customers’ disruption risks, or retention risks. Moreover, we derive the explicit optimal allocation solution under capacity shortage. Interestingly, the optimal allocation does not prioritize customers with higher current order volume, because the customers with high volume cost more capacity to retain. Surprisingly, the optimal allocation quantity is independent of future order volume or disruption risk. Furthermore, loyalty could be prioritized or dis-prioritized, depending on customer’s sensitivity to service level. The independency of optimal capacity allocation over the customers’ future demand volume and disruption risk disappears in the finite horizon setting.

In summary, this dissertation analyzes different aspects of risk management in extended global supply chains, using a variety analytical tools, such as game theory, stochastic processes, stochastic comparisons, optimization, and dynamic programing. Together, these works contribute to modeling, decision tools, insights, and guidelines for effective management of responsible sourcing, customer-base diversification, and strategic safety-stock and capacity planning for random demand surges.
Managing Social Responsibility in Multi-tier Supply Chains

2.1 Introduction

On August 2, 2014, an explosion at a factory owned by the Kunshan Zhongrong Metal Products Co. Ltd. in Kunshan City, China, tragically killed 68 and injured 187 workers. According to a subsequent investigation, the explosion was caused by an exposed flame in a workshop filled with combustible metal dust. The primary business of Zhongrong is to provide aluminum wheel hubs for Dicastal, a key supplier for automotive manufacturer General Motors. For years, the workers of Zhongrong had complained about the metal dust build-up and poor ventilation in the factory; however, General Motors delegated the management of Zhongrong (in Tier 2) to its immediate supplier Dicastal (in Tier 1), and hence did not directly monitor or work to improve the safety of working conditions in Zhongrong’s factory (Shirouzu and Li (2014)). In a different incident, in 2007 an unauthorized Tier 2 supplier in Mattel’s supply chain used paint with an excessive content of lead in the production of children’s toys, prompting a series of recalls that ultimately cost Mattel $110
million and resulted in a severe drop in consumer confidence towards the company (Hoyt et al. (2008)). In both cases, a supplier’s negligence, poor investment in worker safety, or unsafe production practices led to a disastrous outcome, and in both cases, the supplier engaging in such behavior was not the immediate supplier of a large multinational corporation like General Motors or Mattel, but rather a higher tier supplier with whom the downstream firm had no direct relationship. Such events, while tragic, are not uncommon. In a study examining 3,922 supplier relationships, Sedex (2013)\textsuperscript{1} determined that the risk of social and environmental responsibility violations increases in the higher tiers of the supply chain in both number and severity: the average number of responsibility violations is 18\% higher in Tier 2 than in Tier 1, and 27\% higher in Tier 3 than in Tier 2, and a larger fraction of incidents are classified as “major” or “critical” in higher tiers than in lower tiers.

Such violations can result in substantial financial costs to the buying firm, despite the fact that they occurred a significant “distance” away from the buyer (in supply chain terms). On one hand, the downstream firm may experience an immediate loss in demand if socially conscious consumers refuse to purchase products made in an unsustainable, unethical, or unsafe supply chain (Salfino (2014); Guo et al. (2014)). Such losses impact demand for products sourced from the violating supplier, and hence have a magnitude proportional to the volume sourced from the offending supplier; we call such a cost the “direct loss” from a responsibility violation. On the other hand, responsibility violations can have broader ramifications beyond a reduction in the demand for the violating product. For example, the brand equity of the downstream firm can be significantly negatively impacted, causing demand losses for other products not sourced from the violating supplier or forcing the firm to

\textsuperscript{1} Sedex is a not for profit membership online database which allows members to store, share, and report on information in areas like labor, standards health & safety, environment, and business ethics. It is the world’s largest collaborative platform for managing and sharing multi-tier supply chain data, with more than 30,800 members across 23 different industry sectors in 150 countries. Its members includes P&G, Nestle, Unilever, etc.
expend significant resources to recover its reputation with its customer base (Lefevre et al. (2010); Doorey (2011)). We call this cost, which is not necessarily proportional to the volume sourced from the offending supplier, an “indirect loss” resulting from violations in the supply chain.

The potential for responsibility violations at upper tier suppliers leads to a unique challenge for downstream buyers: broadly speaking, should the buyer directly manage the responsibility of its upper tier suppliers, or should it delegate the management of responsibility to its direct (Tier 1) partners? A majority of companies (93% according to Slowick (2013)), including General Motors, Wal-Mart, and other large multi-nationals, choose to work only with their immediate suppliers, delegating control of responsibility efforts in higher tiers to their Tier 1 suppliers. Consequently, in a delegation strategy, the manufacturer (which we also refer to as the “Tier 0 firm”) engages in a relationship with the Tier 1 firm, and the Tier 1 firm monitors and manages the responsibility level of the Tier 2 supplier. Alternatively, some companies, including Hewlett-Packard and Migros, directly manage responsibility at their higher tier suppliers via initiatives such as supplier certification, performing factories audits, and providing a supplier compliance training program (Grimm et al. (2014)). Thus, in a control strategy, while the Tier 0 firm sources a product from Tier 1, it also engages directly with Tier 2 to monitor and improve the level of social and environmental responsibility. Indeed, in theory firms need not follow either extreme strategy, and may instead employ a combination of both strategies, combining direct effort from Tier 0 with effort from a Tier 1 supplier to help manage the responsibility of a Tier 2 supplier.

The advantage of a control strategy is that the downstream firm can directly choose the effort exerted at a Tier 2 supplier to improve responsibility, avoiding the moral hazard problem inherent in a delegation strategy (Kayis et al. (2013)). However, this could be costly, especially when Tier 1 suppliers have closer relationships
with Tier 2 suppliers or more specific expertise or knowledge of the production process than Tier 0 firms. Moreover, even if direct control is cost efficient, the number of suppliers typically increases significantly in higher tiers of the supply chain, implying that a Tier 0 firm pursuing this strategy must manage the responsibility of a large and complex network of suppliers, a potentially complicated task. These observations lead us to pose the following research questions. First, under what conditions should the downstream firm choose the control strategy, the delegation strategy, or some combination of the two? Are there any particular circumstances of the motivating examples discussed above that suggest one strategy or the other should be preferred? Second, what impact does pressure from external stakeholders (consumers, NGOs, and the government) have on the level of supply chain responsibility, both under the manufacturer’s optimal strategy and when the manufacturer is forced to use one particular strategy, e.g., due to supply chain complexity? And third, is there any mechanism by which a decentralized multi-tier supply chain may achieve the first best (centralized optimal) level of responsibility?

We answer these questions by analyzing a model of a stylized, three tier supply chain consisting of one Tier 0 buyer, or “the manufacturer,” one Tier 1 supplier, and one Tier 2 supplier. As our focus is to investigate how the manufacturer should manage responsibility risk at upper tier suppliers, we assume the Tier 1 supplier is perfectly responsible with zero risk of a violation; all responsibility violations come from the Tier 2 supplier. The risk at the Tier 2 supplier is characterized by a Bernoulli responsibility distribution: with some probability, the Tier 2 supplier is “compliant” (meaning it experiences no responsibility violations), and with complementary probability the supplier is “non-compliant,” i.e., experiences a responsibility violation that leads to both direct and indirect losses, as described above. To reduce the probability of a violation, both the Tier 1 supplier and the manufacturer can exert effort at the Tier 2 supplier. Their efforts are substitutable, representing activities
such as certification or factory auditing. While effort is not directly contractable, the manufacturer is powerful and capable of designing its procurement contract with the Tier 1 supplier (e.g., establishing a wholesale price).

We find that either a pure control or a pure delegation strategy is always optimal for the manufacturer, and we discuss conditions under which each strategy is preferred. When the manufacturer is forced to operate using a single strategy—either always working through the Tier 1 supplier or always managing responsibility directly—greater external pressure always increases the supply chain’s responsibility level. However, when the manufacturer has the option to bypass the Tier 1 supplier and work directly with a Tier 2 supplier—in other words, when the manufacturer can optimally choose between a control strategy and a delegation strategy—increased external pressure may generate perverse incentives with the manufacturer, causing the manufacturer to change strategies and reduce the overall investment in responsibility. Specifically, increased pressure from consumers (in the form of larger boycotts from responsibility violations) or NGOs (in the form of more frequent audits and a higher probability of detecting violations) may actually decrease the level of responsibility in the supply chain, because this pressure causes the manufacturer to switch from the control strategy to the delegation strategy, which results in less effort in equilibrium. Greater responsibility effort is guaranteed only when pressure is imposed on the manufacturer in the form of greater indirect losses (e.g., penalties or fines from violations by governmental bodies). We also show that appropriately designed non-compliance penalties charged to either the Tier 1 supplier or the manufacturer whenever the Tier 2 supplier experiences a violation can coordinate the supply chain, achieving the centralized system optimal effort level, and eliminate counterintuitive responses to external pressure by aligning the firm’s actions with the intentions of external stakeholders. However, we discuss several practical issues surrounding the optimal penalty contracts which might make their implementation challenging.
The remainder of this paper is organized as follows. In §2.2, we briefly review the literature. §2.3 introduces the components of our model, §2.4 derives the manufacturer’s optimal strategy, and §2.5 analyzes the impact of external stakeholder on pressure on that strategy. §2.6 discusses how non-compliance penalties influence the results, and §2.7 extends our analysis to cover the case of complementary (rather than substitutable) effort activities by the Tier 1 supplier and manufacturer. §2.8 concludes the paper.

2.2 Literature Review

Supply chain social and environmental responsibility is a topic that has received increasing attention in recent years. The most relevant papers to our own in this stream are those that consider sourcing and supplier management strategies to improve social responsibility. Guo et al. (2014) consider how a buyer chooses between multiple suppliers, some of which are “responsible” and some of which are “non-responsible,” and analyze how external factors (e.g., consumer willingness-to-pay for responsibly sourced products) influences the buyer’s sourcing decision. In contrast to their model, which consists of fixed (exogenous) levels of responsibility, the Tier 2 supplier in our model has an endogenous level of responsibility that is determined by the effort levels of the Tier 0 and Tier 1 firms. While they also find that increased external pressure can lead to lower supply chain responsibility, in their model this occurs because of the buying firm’s desire to exploit consumer willingness-to-pay for responsibility, while consumers in our model have no such willingness-to-pay; rather, our results are driven by the manufacturer’s trade-offs between the direct control and delegation strategies. Xu et al. (2015) consider mechanisms to reduce a direct supplier’s incentive to use child labor in the production of the buyer’s product. In contrast to their model, in which the buying firm is forced to operate through the Tier 1 supplier and endure moral hazard, in our model the buyer has the option
of bypassing the Tier 1 supplying and engaging directly with its Tier 2 supplier to improve social responsibility. In other related works, Kalkanci et al. (2012) compare social and environmental performance under voluntary and mandatory disclosure rules about product responsibility; Lewis et al. (2014) analyze supply agreements for sustainable quality supply under asymmetric information; Plambeck and Taylor (2013) show how to motivate a supplier’s compliance with environmental regulations when the supplier can evade a buyer’s audit; Chen and Lee (2014) compare different screening mechanisms and determine how they motivate the responsibility level of a firm’s immediate supplier; Chen and Lee (2014) discuss how commitments between supply chain members can improve social responsibility, and Kim (2015) analyzes the optimal inspection frequency and non-compliance penalty for a firm that randomly experiences violations. Each of these papers consider asymmetric information regarding violations or responsibility levels, while in our model, all information is public; rather, the key tension in our model derives from the non-contractibility of the Tier 1 supplier’s effort to improve responsibility in Tier 2, combined with the capability of the Tier 0 buyer to directly manage responsibility in Tier 2 if it so desires.

A related topic is the management of risk in decentralized supply chains, specifically the risk deriving from supply chain disruptions in which an upstream supply chain member (e.g., a supplier) takes actions that impact the supply risk of the downstream supply chain member (e.g., a manufacturer). Many different strategies for mitigating the risk of supply chain disruptions have been examined, including multi-sourcing, identifying sources of back-up supply, and carrying safety stocks (Anupindi and Akella (1993); Tomlin (2006); Babich et al. (2007)). See Aydin et al. (2010) for a recent review of the literature in decentralized supply chain disruption risk management. In contrast to these studies, we consider supplier risks from social and environmental responsibility violations, which impacts the downstream firm via
direct losses from reduced demand and indirect financial losses, rather than through a loss in supply; because of this, many of the above mitigation strategies for supply disruptions are not effective (e.g., carrying safety stocks is not an effective mitigation strategy for a disruption in the Tier 0 firm’s demand).

In addition, we consider a multi-tier supply chain, rather than the vast majority of works in the supply chain disruption literature which focus on two-tier (supplier-manufacturer) supply chains. However, there are a few existing works that explore risk management in multi-tier supplier chains. Ang et al. (2014), for instance, examine a supply chain with a manufacturer, Tier 1 suppliers, and risky (prone to disruption) Tier 2 suppliers under different endogenously determined supply chain structures. Bimpikis et al. (2013) also consider a manufacturer’s sourcing decision when Tier 1 suppliers endogenously choose amongst risky Tier 2 suppliers, with an emphasis on the impact of non-convexities in production function on formation of particular supply networks. While these papers assume that the manufacturer must always source through the Tier 1 supplier, we differ by focusing on the downstream buyer’s the choice of either working directly with a Tier 2 supplier or working through a Tier 1 supplier. Dong et al. (2014) discuss a quality control problem in a multi-level supply chain, comparing the optimal policy when the manufacturer does and does not directly source from a Tier 2 component supplier. In contrast to that work, we consider Tier 2 suppliers prone to responsibility risk, which, unlike quality risk, cannot be determined by inspecting the product at the Tier 1 stage, and unlike supply disruption risk, results in direct demand losses and indirect financial losses rather than supply losses.

Lastly, we note that in our model, a buying firm can delegate the management of responsibility risk in Tier 2 to a Tier 1 supplier. The effort exerted by that Tier 1 supplier to improve responsibility in Tier 2 is non-contractable, meaning the relationship between the buyer and Tier 1 supplier is that of a principal and an agent
(Laffont and Martimort (2009)). Hence, our work is also related to the extensive literature on principal-agent models in supply chains (e.g., Corbett and de Groote (2000), Yang et al. (2009)) specifically those involving moral hazard (e.g., Krishnan et al. (2004), Babich and Tang (2012), Hwang et al. (2014)), as the principal in our model (the buying firm) faces risk that is dependent on the actions of the agent (the Tier 1 supplier). In contrast to most of the literature involving principal-agent relationships in supply chains, the multi-tier structure of our model implies that the principal can, at its discretion, completely bypass the Tier 1 agent and directly exert effort to manage the responsibility risk of a Tier 2 supplier. A number of recent works have explored this question in both economics and operations management. For example, Kayis et al. (2013) consider the choice between delegation and control when the decision in question is the procurement of a subcomponent, and Guo et al. (2010) analyze the performance of various outsourcing structures that are feasible in a three tier supply chain; we refer the reader to these works for a detailed summary of this literature. To the best of our knowledge, the trade-off between delegation and control has not been considered in the context of supplier social and environmental responsibility risk.

2.3 Model

We consider a three level serial supply chain with an end company (Tier 0), a Tier 1 supplier, and a Tier 2 supplier. For simplicity, we use “manufacturer” to refer to the end company, which can be either a conventional manufacturer or a consumer-facing retailer. The manufacturer faces one-time deterministic demand $d$ from its customers and sells products at unit price $p$. To provide the products, the manufacturer sources a critical component from the Tier 1 supplier at unit price $w$, and the Tier 1 supplier in turn sources critical components from the Tier 2 supplier at unit price $c$. We assume the selling price $p$ is fixed as might be the case, e.g., if the price
is determined by market competition. In addition, the unit price $c$ from the Tier 2 supplier is exogenous, given that suppliers at this level produce commodity-type components with relatively stable prices. Naturally, we require $c \leq w \leq p$, so that it is profitable for each party to participate in selling the product. All other transaction or administrative costs are normalized to zero.

We assume the Tier 1 supplier is perfectly responsible while the Tier 2 supplier is subject to a risk of social and environmental responsibility violations. Examples of such violations, which we also call “non-compliance incidents,” include illegal and forced labor, factory fires or other accidents that occur due to negligence, or illegal hazardous waste emissions. Consistent with much of the risk management and responsibility literature (e.g., Tomlin (2006) and Guo et al. (2014)), we model Tier 2 supplier responsibility risk with a Bernoulli distribution dictating whether the supplier is “non-compliant” with probability $\alpha(e)$ or “compliant” with probability $1 - \alpha(e)$, where $e$ is the effort exerted by the manufacturer and Tier 1 supplier to improve responsibility in Tier 2. The function $\alpha(e)$ thus determines the likelihood of a violation, while the magnitude of a violation is determined by both the direct and indirect loss parameters (described below). We assume that a responsibility violation occurs if (i) the supplier is non-compliant, and (ii) the non-compliance is detected. The former factor is purely a function of the efforts of the Tier 1 supplier and manufacturer, while the latter (detection of non-compliance) may be a function of, for instance, the frequency of audits by the firm or NGOs, or the amount of media pressure focused on the manufacturer’s supply chain.\footnote{For example, certain companies, such as Apple or Wal-Mart, may draw greater scrutiny from the press and NGOs due to their large global presence and dominance in the market; this greater scrutiny would result in a higher chance of non-compliance being detected, and would be reflected in our model via larger $\beta$ term.} In particular, we adopt the functional form $\alpha(e) = \beta \lambda(e)$, where $\beta \in [0, 1]$ is the detection rate of responsibility violations and $\lambda(e)$ is a decreasing convex function that represents the probability
of the Tier 2 supplier actually experiencing a responsibility violation (determined by Tier 1 supplier and manufacturer effort). Where necessary, we will also discuss the special case of the exponential risk function, $\alpha(x) = \beta e^{-x}$.

To reduce the Tier 2 supplier’s probability of a responsibility violation $\alpha(e)$, both the Tier 1 supplier and the manufacturer can exert effort at the Tier 2 supplier. This may include more frequent audits, a supplier training program to implement responsibility standards, technological support, or direct financial subsidies for, e.g., safety improvements in the supplier’s facilities. We let $e_i$ be the effort the Tier $i$ company exerts, and the cost of effort at firm $i$ is $k_i e_i$, where $k_i > 0$ is a constant.

We consider two cases regarding the form of total effort $e$:

- **Substitutable Effort**: $e = e_0 + e_1$. This is suitable for efforts such as more frequent supplier monitoring and audits, e.g., to prevent child labor or excessive over-time, that are essentially interchangeable. In these situations, the total effort exerted in the supply chain is the sum of efforts from both the Tier 1 supplier and the manufacturer. This case is analyzed in the main portion of the paper.

- **Complementary Effort**: $e = e_1 * e_0$. This is suitable for more complex responsibility improvement activities where the Tier 1 supplier and manufacturer each play an important, and independent, role in the success of the initiative. For instance, the manufacturer may possess the financial resources to purchase new, safer equipment for the Tier 2 supplier, while the Tier 1 supplier possesses the resources to train the Tier 2 supplier in the proper usage of the equipment and to continually monitor its effectiveness. In this case, successful investment requires effort from both companies, and greater effort from one firm makes the effort of the other firm more effective. We consider this case as an extension in §2.7.
As discussed in the introduction, a responsibility violation leads to both direct and indirect loss for the manufacturer. To model the former effect, we assume that a percentage $\gamma$ of the manufacturer’s customers are “socially conscious” and abandon the manufacturer in the wake of a violation, leading to a demand loss $\gamma d$ (Guo et al. (2014)). Hence, the direct effect is proportional to the manufacturer’s demand and will, in turn, affect the quantity sourced from the Tier 1 supplier. To model the latter effect, we assume that a violation leads to an indirect financial loss $a_0$ for the manufacturer (Plambeck and Taylor (2013)). This includes any factors that are not proportional to the manufacturer’s demand and that do not impact the order the manufacturer places with the Tier 1 supplier. For instance, it may derive from recovery costs not directly associated with the non-compliant product, e.g., advertising costs to recover consumer confidence; alternatively, it may derive from other mechanisms such as a reduction in the manufacturer’s stock price (Lefevre et al. (2010)). In addition, such costs may arise if the manufacturer must subsidize the supplier’s recovery efforts or compensate victims of the supplier’s negligence. For instance, after a fire in a Bangladeshi garment factory in 2012 which killed more than 100 workers, Li & Fung and other manufacturers whose products were made in Bangladeshi factories provided compensation for the victims of the fire and their families (?). Finally, the indirect loss term may also be impacted by additional fines or penalties applied by a regulator or governmental body. In addition to the manufacturer, the Tier 1 supplier may encounter a similar indirect financial loss $a_1$, deriving from similar sources. Figure 2.1 illustrates the supply chain structure and the impact of responsibility violations on the firms in our model.\(^3\)

\(^3\) We note that in practice, following a violation the Tier 2 supplier may lose some or all of its operating capacity, e.g., due to factory suspension or shut-down. For example, after the fire in the Bangladeshi apparel factory, not only was the affected factory clearly shut down for some time, many other factories in Bangladesh were also suspended while safety regulations were reviewed and updated. For simplicity and to focus on the impact of social and environmental responsibility risk on demand, we will not explicitly model capacity loss resulting from violations; however, our main
The manufacturer has full information regarding the supply chain, including the responsibility risk function $\alpha(e)$ and the effort cost parameter $k_i$ of the Tier 1 supplier. However, the Tier 1 supplier’s effort $e_1$ is not directly contractable, e.g., because effort is abstract and ill-defined, and verifying effort in a legally enforceable manner is challenging. In practice, a variety of incentives have been adopted by manufacturers to motivate the Tier 1 supplier’s effort. These strategies can be broadly categorized as “inducements” or “penalties.” For example, Starbucks pays 10% more for coffee beans grown with the certification of fair trade, i.e., it induces responsible behavior with a high wholesale price. On the other hand, Wal-Mart and Sam’s Club have adopted a strict penalty policy, which terminates a supplier when its annual non-compliance violations exceeds a certain number. We consider both types of incentives in this paper. In the main analysis in §§2.4-2.5, we assume that the manufacturer offers a simple wholesale price only contract with parameter $w$, i.e., penalties are not allowed. In §2.6, we consider the impact of non-compliance penalties by allowing the manufacturer to charge a fixed penalty $\varphi$ to the Tier 1 supplier whenever a violation occurs in Tier 2. In both cases, the Tier 1 supplier will accept any contract as long as its expected profit from the contract is no lower than its reservation level, $r$.

The manufacturer is assumed to be powerful (e.g., consistent with our motivating examples of firms such as General Motors and Mattel) and serves as a sequential leader both in setting the terms of the contract with Tier 1 and in determining its results continue to hold after incorporating capacity losses driven by responsibility violations.
own responsibility efforts. The sequence of events is thus divided into three stages—a contracting stage, a production stage, and a selling stage—as follows. (1) At the beginning of the contracting stage, the manufacturer determines its effort level $e_0$ and chooses a contract to offer the Tier 1 supplier. After observing the contract and the manufacturer’s effort level, the Tier 1 supplier decides to accept or reject the offer. If it accepts the offer, it determines its effort $e_1$ to maximize its profit. (2) During the production stage, the efforts $e_1$ and $e_0$ take effect. Responsibility violations then randomly occur with probability $\alpha(e)$, and the impact of a violation (demand loss $\gamma$ and financial loss $a_i$) is realized. (3) After observing the non-compliance state, the manufacturer chooses its order quantity $q$ from the Tier 1 supplier. Sales occur and the profit for each company is calculated.\footnote{Note that in the selling stage, the manufacturer finalizes its order quantity after violations have been revealed; this is reasonable, because the Tier 2 supplier takes responsibility for non-compliance and in many cases would allow the manufacturer to adjust its order quantity $q$ should non-compliance occur (see also Guo et al. (2014)); moreover, if there is any capacity loss in Tier 2 associated with a responsibility violation (e.g., due to a factory shutdown in the wake of an accident or regulatory investigation), a natural opportunity exists for the manufacturer to adjust its order downward to match the reduced demand level.} Figure 2.2 provides an illustration of the sequence of events.

The objective of each player is to maximize its individual expected profit. For a general contract consisting of both a wholesale price $w$ and a non-compliance penalty $\varphi$, let $\Pi_i(w, \varphi, e_0, e_1)$ denote the expected profit function for the Tier $i$ firm. Then
the manufacturer solves the following optimization problem:

$$\max_{p \geq w \geq c, e_0, \varphi \geq 0} \Pi_0 (w, \varphi, e_0, e_1^*(w, \varphi, e_0))$$

s.t. $$e_1^*(w, \varphi, e_0) = \arg\max\{e_1 \geq 0 : \Pi_1 (w, \varphi, e_0, e_1)\}$$

$$\Pi_1 (w, \varphi, e_0, e_1^*(w, \varphi, e_0)) \geq r$$

We assume the problem is feasible, i.e., the model parameters are chosen so that there exists some \((w, \varphi, e_0)\) so that the Tier 1 supplier has profit larger than its reservation level \(r\). Because demand is deterministic, the manufacturer always finds it optimal to order its demand, leading to \(d\) under a compliance state and \((1 - \gamma) d\) under a non-compliance state. We may thus write the expected profit functions as

$$\Pi_0 (w, \varphi, e_0, e_1) = (p - w) d - k_0 e_0 - \alpha (e) L_0 (w, \varphi)$$

$$\Pi_1 (w, \varphi, e_0, e_1) = (w - c) d - k_1 e_1 - \alpha (e) L_1 (w, \varphi)$$

where \(L_1 (w, \varphi) = \gamma (w - c) d + a_1 + \varphi\) and \(L_0 (w, \varphi) = \gamma (p - w) d + a_0 - \varphi\). Here, \(L_i (w, \varphi)\) represents the Tier \(i\) company’s profit loss when a responsibility violation occurs in Tier 2. It includes the direct loss due to losing \(\gamma d\) demand, the indirect financial loss \(a_i\), and the internal contract penalty term \(\varphi\).

### 2.4 The Manufacturer’s Optimal Strategy

We begin by investigating the manufacturer’s optimal strategy under a wholesale price contract with no non-compliance penalty (\(\varphi = 0\)). Such contracts are widely observed in industry, due to their simplicity and convenience for negotiation and legal implementation (Kayis et al. (2013)), especially when suppliers are averse to contractible liability for upper tier supply chain responsibility violations. We first establish a baseline for comparison by deriving the optimal strategy in a centralized supply chain, i.e., a supply chain in which a single decision maker determines the effort levels of both the Tier 1 supplier and the manufacturer in order to maximize
total supply chain profit. The expected total supply chain profit can be written as

$$\Pi(w, \varphi, e_0, e_1) = (p - c) d - k_0 e_0 - k_1 e_1 - \alpha(e) L$$

Here, $L = \gamma (p - c) d + a_1 + a_0$ is the supply chain profit loss when a responsibility violation occurs at the Tier 2 supplier. Optimizing this profit function over manufacturer and Tier 1 effort yields the following result:

**Proposition 1.** In a centralized supply chain, it is optimal for the firm with lower effort cost to exert all effort, yielding

$$e^* = \begin{cases} 0 & \text{if } \min_{k_0, k_1} \frac{L}{L} \geq -\alpha'(0), \\ (\alpha')^{-1} \left( -\min_{k_0, k_1} \frac{L}{L} \right) & \text{otherwise.} \end{cases}$$

Recall that substitutable effort activities, such as supplier factory audits or compliance monitoring, can be conducted by either Tier 0 or Tier 1 with no difference in impact on the risk of a violation. Therefore, the optimal strategy is intuitive: the system only exerts effort when the effort cost is low compared with the supply chain loss, and all effort should come from the firm with lower effort cost. Since Tier 1 suppliers usually have better information about and access to Tier 2 suppliers, a reasonable scenario in practice is that the Tier 1 supplier has lower unit effort cost than the manufacturer. In this case, the proposition shows that it is better to let the Tier 1 supplier handle all responsibility improvement efforts. However, it is also possible in practice that the manufacturer has greater knowledge about responsibility practices or more effective implementation tools, leading to $k_0 < k_1$; in these cases, the centralized optimal solution is for the manufacturer to exert all effort. To keep our focus on interesting scenarios, we assume it is optimal to exert non-zero effort under the centralized setting in the remainder of our paper. Specifically, we assume $k_i \leq -\alpha'(0) a_i$ under substitutable effort, which means the unit investment cost is
lower than the indirect responsibility violation loss, which in turn ensures $e^* > 0$ is optimal in a centralized system.

Next, we consider a decentralized supply chain in which the manufacturer is forced to always follow one extreme strategy or the other, i.e., either delegation or control, but cannot optimally choose between the two. This might be the case if, for instance, Tier 1 suppliers do not have the means or resources to improve responsibility in Tier 2 (forcing the manufacturer to use a control strategy) or if the manufacturer is reluctant to become involved in the management of higher tier suppliers due to a lack of information about Tier 2 or the complexity of managing what in practice is a potentially large number of Tier 2 suppliers (forcing a delegation strategy). Define $e_0^c (w) (e_1^d (w))$ to be the optimal effort level of the investing party in a pure control (pure delegation) strategy for some wholesale price $w$; from this point forward, we use the superscript $c$ to denote optimal values under a control strategy and $d$ for optimal values in a delegation strategy. Then, we have:

$$e_0^c (w) = \left( \alpha \right)^{-1} \left( -\frac{k_0}{L_0(w,0)} \right) \quad \text{and} \quad e_1^d (w) = \left( \alpha \right)^{-1} \left( -\frac{k_1}{L_1(w,0)} \right)$$

(2.1)

In other words, $e_0^c (w)$ is the optimal manufacturer effort when Tier 1 effort is fixed to zero ($e_1 = 0$), while $e_1^d (w)$ is the optimal Tier 1 effort when manufacturer effort is fixed to zero ($e_0 = 0$). In both cases, the more costly the firm’s effort relative to its non-compliance loss, the lower effort that firm exerts. The manufacturer thus acts as a sequential leader in setting the wholesale price $w$ to maximize its own profit, knowing that the induced effort level will be given by one of the two equations given in (2.1). We examine how these effort levels depend on the problem parameters after optimizing over the wholesale price $w$ in §2.5.

Lastly, we move to a decentralized supply chain in which both the manufacturer and Tier 1 supplier independently determine their effort levels, and simultaneous positive effort from both parties is feasible. We begin by deriving the equilibrium
effort levels given some wholesale price \( w \), then consider the subsequent equilibrium effort strategy under the optimal choice of \( w \) by the manufacturer. Let \((e_0^*(w), e_1^*(w))\) be the equilibrium effort levels induced by a particular wholesale price. Note that the manufacturer’s effort, \( e_0^*(w) \), is chosen first, and the Tier 1 supplier’s effort, \( e_1^*(w) \), is chosen second as a best response to the manufacturer’s effort and the wholesale price. Hence, the manufacturer induces some \( e_1^*(w) \) given its own choice of effort level. We thus refer to \((e_0^*(w), e_1^*(w))\) as the manufacturer’s optimal induced effort strategy given a wholesale price \( w \). Given these definitions, we have the following result:

**Proposition 2.** In a decentralized supply chain with a wholesale price contract, there exists a threshold price \( c \leq w_s \leq p \) such that the manufacturer’s optimal induced effort strategy is

\[
(e_0^*(w), e_1^*(w)) = \begin{cases} (e_0^*(w), 0) & , w \leq w_s \\ (0, e_1^*(w)) & , w > w_s \end{cases}
\]

The proposition shows that the optimal induced effort strategy of the manufacturer for a fixed wholesale price is either a pure control strategy (for a low wholesale price) or a pure delegation strategy (for a high wholesale price), mimicking the effort levels derived in (2.1). No intermediate strategy, in which both firms exert positive effort, can be optimal. Thus, consistent with the centralized system, only one party exerts effort at the equilibrium; however, which party takes that responsibility depends on the wholesale price \( w \). The fact that such “pure” strategies are optimal occurs directly because of the substitutable nature of responsibility improvement efforts. If the wholesale price is low, then the Tier 1 supplier’s incentive to exert effort is small, meaning the manufacturer must exert all effort by following a control strategy; conversely, if \( w_s \) is large, the Tier 1 supplier has strong incentive to exert effort, and hence the manufacturer will exert zero effort and follow a delegation strategy.
Interestingly, while we might expect the manufacturer to follow a control strategy whenever it has a larger incentive to exert effort \((a_0 > a_1)\) and its effort is more cost efficient than the Tier 1 supplier’s effort \((k_0 < k_1)\), Proposition 2 implies that this is not necessarily true; in fact, the manufacturer’s strategy preference depends on the wholesale price, and even if \(k_0 < k_1\) and \(a_0 > a_1\), it is possible for some wholesale price that the manufacturer would prefer to follow a delegation strategy. Moreover, note that the Tier 1 supplier’s reservation profit level \(r\) is often interpreted as a proxy for the bargaining power of the supplier. When the Tier 1 supplier is very powerful (has a high reservation profit), this may lead to a high wholesale price (above \(w_s\)), which in turn would push the manufacturer to pursue a delegation strategy even if it is more efficient than the Tier 1 supplier in improving Tier 2 supplier responsibility.

Next, we analyze the manufacturer’s preference between these two extreme strategies under the optimal wholesale price \(w^*\). Let \(\Pi^s_i(w)\) denote the expected profit of firm \(i\) under strategy \(s \in \{c, d\}\), given the optimal effort described in the above proposition. When the manufacturer chooses to manage responsibility directly, its profit function is decreasing in \(w\) and thus it has no incentive to share more than the minimal reservation profit with the Tier 1 supplier. As a result, the optimal wholesale price is \(w^c\), where \(w^c\) is the minimum price satisfying \(\Pi^c_i(w) = r\). By contrast, if the manufacturer chooses to delegate responsibility management to the Tier 1 supplier, it needs to offer a wholesale price no lower than \(\hat{w}\), where \(\hat{w}\) is the minimal price satisfying \(\Pi^d_i(w) \geq r\). Thus, when the control strategy is optimal, the manufacturer extracts the maximal unit profit by setting \(w^* = w^c\) and exerts effort \(e^*_0 = e^c_0(w^c)\). On the other hand, when the delegation strategy is optimal, the manufacturer incentivizes the Tier 1 supplier to exert effort by offering a wholesale price \(w^* \geq \hat{w}\). Given these observations, we can identify under what circumstances each strategy is preferred:
Proposition 3. In a decentralized supply chain under the manufacturer’s optimal wholesale price contract, fixing all other parameters, there exist some thresholds (denoted by the over line \(\bar{\cdot}\) symbol) such that the control strategy is optimal for the manufacturer if and only if:

(i) The cost of effort in Tier 0 is low \((k_0 \leq \bar{k}_0)\);

(ii) The indirect loss in Tier 0 is high \((a_0 \geq \bar{a}_0)\);

(iii) The cost of effort in Tier 1 is high \((k_1 \geq \bar{k}_1)\);

(iv) The indirect loss in Tier 1 is low \((a_1 \leq \bar{a}_1)\);

(v) The reservation profit of the Tier 1 supplier is low \((r \leq \tau)\).

The first four cases are intuitive. In cases (i) and (ii), the manufacturer has strong incentive to exert high effort, either due to low cost \((k_0)\) or high financial cost of a violation \((a_0)\), and hence the manufacturer would prefer direct control to maximize the amount of effort exerted to improve responsibility. Conversely, in cases (iii) and (iv) the Tier 1 supplier does not exert sufficient effort under delegation, due to either high cost \((k_1)\) or low financial costs of a violation \((a_1)\). To understand case (v), recall that the manufacturer sets a very low wholesale price in a control strategy, leaving the Tier 1 supplier with exactly its reservation profit; when using delegation, however, the manufacturer must set a higher price to induce sufficient responsibility effort. Thus, when the reservation profit is very low, the temptation to use the control strategy is strongest, as this strategy allows the manufacturer to extract significant surplus. Thus, when the Tier 1 supplier is “weak,” implying a small \(r\), the manufacturer prefers a control strategy. Put another way, case (v) shows that Tier 2 responsibility management should be left to the more powerful company, i.e., the control strategy is optimal for the manufacturer if the Tier 1 supplier is not powerful \((r\) is small) and the delegation strategy is optimal if the Tier 1 supplier is powerful \((r\) is large). Figure 2.3 illustrates these results by graphically depicting
the manufacturer’s optimal strategy along with the corresponding optimal wholesale price as a function of several different parameter values. In the left panel, which depicts the optimal strategy as a function of $k_0$, $a_1$, or $r$, control is optimal for low parameter values and delegation is optimal for high parameter values. In the middle panel, delegation is optimal for low $k_1$ and control is optimal for high $k_1$. In the right panel, delegation is optimal for low $a_0$ while control is optimal for high $a_0$.

Note that Proposition 3 does not address the impact of direct losses ($\gamma$) or the detection probability ($\beta$) on the manufacturer’s optimal strategy. Intuitively, we might expect when the direct demand loss $\gamma$ increases, the manufacturer should choose the control strategy and more actively manage supply chain responsibility. However, this need not be the case, because unlike $k_i$ and $a_i$, the impact of demand loss $\gamma$ falls on both the manufacturer and the Tier 1 supplier, as demand loss equates to lost production volume for both firms in the supply chain. Hence, the Tier 1 supplier also feels greater incentive to exert effort as $\gamma$ increases, which may cause the manufacturer to shift from a control strategy to a delegation strategy once the supplier has sufficient incentive to improve responsibility. This can lead to the manufacturer’s optimal strategy being non-monotonic in $\gamma$, and consequently, the impact of direct losses on the manufacturer’s optimal strategy can be complex. In a similar manner, the detection probability ($\beta$) can potentially have a non-monotonic impact on the
manufacturer’s optimal strategy.

Lastly, we consider how the decentralized supply chain’s total effort level compares to the centralized supply chain benchmark derived in Proposition 1. From the preceding results it is straightforward to see that:

**Corollary 4.** *Total effort in a decentralized supply chain is lower than total effort in a centralized supply chain.*

Thus, unsurprisingly, the wholesale price contract can never coordinate the supply chain. Much like the “double marginalization” effect commonly observed in decentralized supply chains (Cachon and Lariviere (2005)), where dividing profit among the manufacturer and supplier results in too little inventory, in our model dividing the supply chain’s non-compliance loss $L$ among the manufacturer and Tier 1 supplier leads to responsibility effort under-investment, regardless of whether the manufacturer follows a direct control strategy or a delegation strategy when managing Tier 2 responsibility.

### 2.5 External Stakeholder Pressure

External stakeholders—in particular, socially conscious consumers, non-governmental organizations (NGOs), and policymakers in government—actively seek to encourage companies to exert more effort on improving supply chain responsibility. Socially conscious consumers, for example, influence $\gamma$, the demand loss that occurs in the wake of a responsibility violation, by boycotting products from affected companies. NGOs work with the media and third party auditing groups to increase the detection rate $\beta$ of violating suppliers. Policymakers in government may institute fines on both suppliers ($a_1$) and manufacturers ($a_0$) should a responsibility violation occur. In this section, we examine the impact of each of these external stakeholder actions on the level of responsibility in the supply chain.
Analyzing the effect of consumers and policymakers on responsibility effort can be accomplished via manipulation of the parameters listed above. In all cases, we assume that external stakeholders move first and determine their efforts ($\gamma$, $\beta$, $a_0$, and $a_1$), after which the manufacturer optimizes its effort level and contract with the Tier 1 supplier as in the analysis in the previous section. Throughout this section, we assume an exponential risk function, $\alpha(x) = \beta e^{-x}$, for concreteness, although many of the following results extend to a general risk function.

We begin by examining the impact of external stakeholder actions in a centralized supply chain. As the following proposition shows, in such a situation, as expected, external stakeholder pressure unequivocally increases the level of supply chain responsibility:

**Proposition 5.** In a centralized supply chain, total effort $e^*$ is increasing in consumer pressure ($\gamma$), NGO pressure ($\beta$), and policymaker pressure ($a_0$ and $a_1$).

The reason for this result is that greater pressure along any of the listed dimensions increases the cost of a responsibility violation to the entire supply chain, which therefore increases the optimal level of effort for the centralized system. Most global supply chains, however, are unlikely to be vertically integrated. We thus move to the case of a decentralized multi-tier supply chain analyzed in §2.4. First, we consider the impact of external stakeholder pressure when the manufacturer is forced to use one strategy or the other:

**Proposition 6.** In a decentralized supply chain with a wholesale price contract,

(i) Fixing the manufacturer’s strategy to delegation, total effort $e^*$ is increasing in consumer pressure ($\gamma$), NGO pressure ($\beta$), and policymaker pressure ($a_0$ and $a_1$).

(ii) Fixing the manufacturer’s strategy to control, total effort $e^*$ is increasing in consumer pressure ($\gamma$), NGO pressure ($\beta$), and policymaker pressure on the manufacturer ($a_0$) and decreasing in policymaker pressure on the Tier 1 supplier ($a_1$).
This result shows that if the manufacturer is forced to always use one strategy, greater pressure from external stakeholders continues to unambiguously increase the level of supply chain responsibility in all cases except for one: when the manufacturer engages in direct control and increased pressure falls on the Tier 1 supplier in the form of a greater indirect loss \((a_1)\). The reason for this result is that the manufacturer must compensate the Tier 1 supplier with a greater wholesale price when \(a_1\) increases, in order for the supplier to achieve its reservation profit level \(r\); as a result of this higher wholesale price, the manufacturer earns less margin and thus feels less incentive to exert direct effort to improve responsibility in Tier 2. Nevertheless, in all other cases, external pressure leads to outcomes consistent with the centralized supply chain case. In many practical settings the manufacturer may be forced into one strategy or the other; most models of two-tier supply chains make precisely this assumption, and the complexity and sheer number of Tier 2 (and higher) suppliers may necessitate the use of delegation. In such cases, most sources of external pressure from continue to improve responsibility effort in the supply chain.

In fact, a key feature of a multi-tier structure is that the manufacturer possesses the option of choosing its strategy optimally: delegation or control can be used in accordance with the optimality conditions discussed in Proposition 3. As the following proposition shows, when this is the case, external stakeholders may have quite a different impact on the responsibility effort in the supply chain:

**Proposition 7.** In a decentralized supply chain with a wholesale price contract, under the manufacturer’s optimal strategy, total effort \(e^*\) is increasing in policymaker pressure on the manufacturer \((a_0)\), but may be increasing or decreasing in consumer pressure \((\gamma)\), NGO pressure \((\beta)\), and policymaker pressure on the Tier 1 supplier \((a_1)\).

Proposition 7 shows that when increased policymaker pressure falls only on the
greater external stakeholder pressure leading to lower responsibility effort.

manufacturer, i.e., $a_0$ is larger, the equilibrium system effort increases; in this case, the manufacturer alleviates its loss either by exerting more effort itself or offering a higher price to the Tier 1 supplier to induce greater effort. However, when pressure falls on the entire supply chain (via a higher detection rate $\beta$ or greater consumer pressure $\gamma$), or when pressure falls only on the Tier 1 supplier (a higher Tier 1 supplier penalty $a_1$), equilibrium supply chain effort may be decreasing in external stakeholder pressure.

Figure 2.4 illustrates two examples of this effect.\(^5\) Focusing on the left panel, observe that for low $\beta$, the manufacturer engages in the direct control strategy and exerts all responsibility effort on its own, in the process taking all supply chain profit above the supplier’s reservation level. However, as pressure from the NGO increases, the Tier 1 supplier feels greater loss from a violation, leading it to exert greater effort under a delegation strategy; hence, once $\beta$ is sufficiently large (greater than approximately 50%), the manufacturer shifts its strategy, delegating responsibility management to the Tier 1 supplier. At this point, the total system effort drops

\(^5\) In the figure, the risk function is $\alpha(x) = \beta e^{-x}$ and parameter values are $c = 0.5, p = 3, d = 1, a_0 = 2, k_0 = 0.5, \gamma = 0.5$ and $r = 0$. In the left panel, $a_1 = 1$, while in the right panel, $\beta = 1$. 

30
significantly, because the Tier 1 supplier has both a smaller share of the supply chain’s profit and a smaller indirect loss than the manufacturer. While this shift to delegation increases the risk of a violation for the manufacturer compared to the control strategy, it means the manufacturer does not need to exert any effort, and overall results in greater profit for the manufacturer. As $\beta$ continues to increase and both the supplier and manufacturer feel a greater expected loss from a violation, the supplier exerts greater effort, eventually raising total system effort above the level exerted by the manufacturer under the control strategy. In addition, changes in $\gamma$ can result in a similar impact on total system effort as changes in $\beta$.

On the other hand, when policymakers impose a higher violation penalty $a_1$ for the Tier 1 supplier, total responsibility effort does not necessarily behave monotonically, as the right panel in Figure 2.4 illustrates. When $a_1$ is small, the manufacturer uses the direct control strategy; as $a_1$ increases, the Tier 1 supplier has more incentive to exert effort on its own, and hence the manufacturer switches to a delegation strategy, which reduces total supply chain effort. Note that when $a_1$ is small and the manufacturer chooses a control strategy, greater $a_1$ may lower the supply chain’s total effort level even if there is no shift in the optimal strategy, consistent with Proposition 6 part (ii).

These results are summarized in Table 2.1. As the table demonstrates, when the supply chain is decentralized, external pressure need not improve supply chain

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Centralized Supply Chain</th>
<th>Decentralized Delegation</th>
<th>Decentralized Control</th>
<th>Decentralized Optimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$</td>
<td>$\uparrow$</td>
<td></td>
<td>$\uparrow$</td>
<td>$\uparrow$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$\uparrow$</td>
<td></td>
<td>$\downarrow$</td>
<td>$\downarrow$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>$\uparrow$</td>
<td>$\uparrow$</td>
<td>$\uparrow$</td>
<td>$\uparrow$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>$\uparrow$</td>
<td>$\uparrow$</td>
<td>$\uparrow$</td>
<td>$\uparrow$</td>
</tr>
</tbody>
</table>
responsibility. This effect arises because of the multi-tier structure of our model, and the manufacturer’s option to choose either a control or delegation strategy; whenever the manufacturer is forced to use one strategy, external pressure is far less likely to have detrimental impact on supply chain responsibility. We stress that these counterintuitive results need not always occur; they are only present when the manufacturer’s optimal strategy shifts from control to delegation in response to an increase in $\gamma$, $\beta$, or $a_1$. Such a shift may or may not occur over the feasible range of these parameter values. Nevertheless, these examples illustrate the perverse impact that greater external pressure can have when a firm can optimally choose between delegation and control when managing supply chain responsibility.

2.6 Penalty Contracts

Penalty contracts are often effective for inducing supply chain coordination in moral hazard settings prone to some type of risk (e.g., Ang et al. (2014)). Since our analysis thus far shows that wholesale price contracts always result in under-investment in responsibility effort, one natural question to ask is: can an appropriately designed penalty contract coordinate a decentralized supply chain, and if so, how does this impact the effects of external stakeholders on supply chain responsibility?

Consider first the case when $k_0 \geq k_1$, i.e., the Tier 1 supplier is more cost efficient than the manufacturer. Recall that in this case, under the centralized supply chain it is optimal for all effort to come from the Tier 1 supplier, i.e., $e^*_1 = (\alpha')^{-1} \left(-\frac{k_1}{\xi}\right)$. In order to coordinate the supply chain, the manufacturer should be able to induce this exact effort level from the Tier 1 supplier, while exerting no effort on its own, i.e., $e^*_0 = 0$. Penalties levied on the Tier 1 supplier potentially give that supplier larger incentive to exert effort to reduce responsibility risk; however, a penalty also reduces the Tier 1 supplier’s profit (all else being equal), and hence the supplier’s
willingness to participate in the contract. To compensate for this, the manufacturer needs to offer a higher wholesale price $w$. Thus, the tradeoff between the wholesale price $w$ and penalty $\phi$ results in a more complicated contract optimization problem for the manufacturer. To ensure that the manufacturer wishes to participate in the decentralized supply chain, we must assume its profit under the centralized setting is non-negative, i.e.,

$$\theta^* = \frac{\Pi^* - r}{(p - c) d} \geq 0.$$ 

Here, $\theta^*$ represents the manufacturer’s maximum profit percentage; the higher $\theta^*$ is, the more profit can be allocated to the manufacturer (after accounting for the Tier 1 supplier’s reservation profit). Given this definition, the following proposition describes the coordinating penalty contract:

**Proposition 8.** When $k_0 \geq k_1$, the decentralized supply chain can be coordinated by a penalty contract with the manufacturer attaining its maximum profit $\Pi^* - r$. The coordinating contract has wholesale price $w^* = (1 - \theta^*) (p - c) + c$ and penalty $\phi^* = \gamma (p - w^*) d + a_0$.

Note that this coordinating contract possesses a simple structure, assigning the manufacturer exactly $\theta^*$ percent of the total profit margin $(p - c)$, which greatly facilitates contract implementation. Moreover, given that the optimal penalty term $\phi^*$ transfers all the responsibility risk to the Tier 1 supplier, the manufacturer has the same profit under both a compliant and non-compliant state, i.e., the manufacturer’s profit is no longer stochastic under the optimal penalty contract.

Table 2.2 illustrates how each of the key problem parameters influences the manufacturer’s optimal wholesale price and non-compliance penalty. Interestingly, the optimal wholesale price and penalty do not necessarily move together monotonically, as is superficially suggested by the expressions in Proposition 8. On one hand, the
penalty itself must reflect the manufacturer’s lost profit during a violation, which in turn is smaller under a higher wholesale price, implying a higher wholesale price should be accompanied by a smaller penalty. Hence, if either $k_1$ or $a_1$ increase and the manufacturer’s optimal profit share ($\theta^*$) decreases, leading to a higher wholesale price, the manufacturer will adjust by offering a smaller penalty. On the other hand, if either $a_0$ or $\gamma$ increase, the manufacturer’s loss during a violation increases, necessitating a larger penalty; to compensate for this larger penalty, the manufacturer must offer larger wholesale price to maintain the supplier’s reservation profit $r$. Hence, the wholesale price and penalty may exhibit a complex relationship.

Conversely, when $k_0 < k_1$, the centralized optimal action is for the manufacturer to exert all effort. In this case, any coordinating contract must give the manufacturer proper incentive to induce the system-wide optimal effort level; since the manufacturer bears all of the cost of effort, this can only be achieved if the manufacturer also feels all of the loss associated with responsibility violations. Hence, the optimal coordinating contract will include a penalty on the manufacturer that transfers all of the Tier 1 supplier’s loss during a violation to the manufacturer. Using the same contract variables as in the above analysis, this equates to a negative $\varphi$ parameter, as the following proposition shows:

**Proposition 9.** When $k_0 < k_1$, the decentralized supply chain can be coordinated by a penalty contract with the manufacturer attaining its maximum profit $\Pi^* - r$. The coordinating contract has wholesale price $w^* = r/d + c$ and penalty $\varphi^* = -\gamma r - a_1$.  

---

**Table 2.2: Impact of model parameters on the coordinating contract ($w^*, \varphi^*$)**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Wholesale Price $w^*$</th>
<th>Penalty $\varphi^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\uparrow k_1$</td>
<td>$\uparrow$</td>
<td>$\downarrow$</td>
</tr>
<tr>
<td>$\uparrow a_1$</td>
<td>$\uparrow$</td>
<td>$\downarrow$</td>
</tr>
<tr>
<td>$\uparrow a_0$</td>
<td>$\uparrow$</td>
<td>$\downarrow$ if $\gamma &lt; \frac{a_0}{\alpha^*}$, $\uparrow$ otherwise</td>
</tr>
<tr>
<td>$\uparrow \gamma$</td>
<td>$\uparrow$</td>
<td></td>
</tr>
</tbody>
</table>

34
This shows that in order to achieve the first best outcome in a decentralized supply chain when the manufacturer is more efficient at responsibility effort, the manufacturer must tell the Tier 1 supplier that (i) it will take full responsibility for engaging the Tier 2 supplier in responsibility improvement, and (ii) the manufacturer will compensate the Tier 1 supplier for its losses in the event of a violation in Tier 2. While intuitive—such a transfer of losses is necessary to place the full level of risk in the hands of the party exerting the effort—such contracts seem far less common in practice than either wholesale price only contracts or “standard” penalty contracts in which the Tier 1 supplier is penalized for violations (perhaps this rarity is due in part to the fact that in most realistic scenarios, the Tier 1 supplier is more efficient at improving responsibility than the manufacturer). Nevertheless, the result shows that the manufacturer can achieve coordination by paying the Tier 1 supplier a low wholesale price and a subsidy whenever a violation occurs, leaving the supplier with exactly its reservation profit $r$ in both the non-compliant and compliant states.

In either of these cases, it is immediately clear that Proposition 5—which details the impact of external pressure on the optimal effort level in a centralized supply chain—holds under the optimal coordinating contract. That is, when the supply chain is appropriately coordinated using penalties, greater external stakeholder pressure always leads to an increase in responsibility effort. This happens because the optimal penalty transfers the entire cost of a responsibility violation to the party exerting effort, and hence any increase in cost due to external stakeholder pressure leads to an increase in effort to improve responsibility. Put another way, under the optimal coordinating penalty contract, the more efficient firm always engages in responsibility improvement efforts, hence changes in external stakeholder pressure do not change the manufacturer’s optimal strategy from control to delegation. Thus, the counterintuitive response to external stakeholder pressure, in which a manufacturer switches strategies and reduces investment in responsibility, is eliminated if the
supply chain properly employs non-compliance penalties.

The use of penalties to transfer supply chain loss from one party to another in order to induce optimal investment is well known in the literature.\(^6\) However, despite the theoretical effectiveness of such contracts, there are key obstacles to their practical implementation (see, e.g., Chen and Lee (2014) and Plambeck and Taylor (2013)). In our setting, several specific challenges arise. First, in the \(k_0 \geq k_1\) case, the Tier 1 supplier may be unwilling to accept a penalty contract. In most practical cases, the dollar amount of the brand reputation damage for the manufacturer is much larger than than that of the supplier, resulting in \(a_0 \gg a_1\). This leads to a high wholesale price \(w^*\), but also a very high penalty \(\phi^*\). Thus, the Tier 1 supplier may be faced with an enormous penalty when a responsibility violation occurs, something that may preclude accepting the contract if the Tier 1 supplier is risk averse or very powerful and capable of dictating whether penalties are employed. Even if the Tier 1 supplier is willing to accept the contract, enforcing the penalty may be difficult, given that many Tier 1 suppliers are small companies from emerging economies whose pockets may not be deep enough to cover such significant penalties. Conversely, if the manufacturer is risk averse, it may also be unwilling to pursue the optimal contract in the \(k_0 < k_1\) case, as this exposes the manufacturer to a potentially large variance in profit. Hence, risk preferences and capital constraints may make implementation of the optimal coordinating contracts challenging. Finally, the Tier 1 supplier may not have incentive to provide accurate information to the manufacturer, such as its effort cost \(k_1\). Note that a higher effort cost leads to a higher wholesale price and lower penalty for the Tier 1 supplier, both favorable outcomes; hence, the Tier 1 supplier may have incentive to lie about or misrepresent this information.

\(^6\) For instance, in contract law under expectation damages, the breaching party is ordered to reimburse the injured party for its losses, which exactly equals \(\phi^*\) in our case. In the operations literature, Plambeck and Taylor (2007) analyze the impact of such damages in repeated relationships and contract design.
in an attempt to generate greater profit. Consequently, while penalty contracts are theoretically effective at eliminating the coordination problem in a decentralized supply chain, these barriers may make their implementation costly, or even infeasible, in practical settings.

2.7 Complementary Effort

In our main model, we assumed that manufacturer and Tier 1 supplier efforts were completely substitutable. This resulted in an optimal allocation of effort that was extreme: in both the centralized and decentralized supply chains, the optimal strategy was either one of pure delegation (Tier 1 exerts all effort) or pure control (the manufacturer exerts all effort). In this section we consider the how our results change under complementary efforts, such as the manufacturer providing equipment and the Tier 1 supplier working with and supervising the Tier 2 supplier in the use of that equipment. Specifically, in contrast to the additive aggregate effort function used in the preceding analysis ($e = e_0 + e_1$), under complementary effort we assume that total effort is the product of the manufacturer’s and supplier’s effort, i.e., $e = e_0 e_1$. For ease of analysis, throughout our discussion of complementary effort we consider the exponential risk function $\alpha(x) = \beta e^{-x}$, although most results are readily extended to a general risk function.

The details of our analysis for this case and the derivation of the manufacturer’s optimal sourcing strategy are contained in Appendix. To summarize, it is clearly no longer optimal for the supply chain to follow the pure delegation or pure control strategies, as under either extreme strategy—in which all effort is exerted by a single firm—aggregate effort is zero. Thus, in both the centralized and decentralized systems, the optimal effort is either zero for both firms or positive for both firms. This fact leads to the following proposition, which illustrates the impact of complementary effort on our key result: the effect of external stakeholder pressure on supply
chain responsibility.

**Proposition 10.** Under complementary effort, in either a centralized supply chain or a decentralized supply chain with a wholesale price contract, the system total effort \( e^* \) is increasing in consumer pressure \( (\gamma) \), NGO pressure \( (\beta) \), and policymaker pressure \( (a_0 \text{ and } a_1) \).

Thus, in contrast to the case of substitutable activities (Propositions 5–7), when responsibility improvement efforts are complementary, external stakeholder pressure has an unequivocal impact on supply chain responsibility: greater pressure from any source always leads to more responsibility effort. This holds regardless of whether the supply chain is centralized, decentralized and coordinated, or decentralized and uncoordinated (i.e., decentralized and operating with wholesale price contracts with no non-compliance penalties). This result arises directly because of the complementary nature of effort and the fact that in equilibrium both firms exert positive (or zero) effort simultaneously. Greater pressure from any external stakeholder leads to at least one member of the supply chain raising its effort level, e.g., greater policymaker pressure on the Tier 1 supplier leads to that supplier raising its effort level. When this happens, the other supply chain member will respond by adjusting its own effort level such that the aggregate effort level \( (e = e_1e_2) \) is raised to a level at least as high as was optimal before the increased pressure, as that firm can accomplish this at a lower cost (i.e., with less of its own effort) than it previously could. Hence, the aggregate effort level will increase in response to greater external pressure.

Thus, when the supply chain responsibility efforts are complementary in nature, more pressure from external parties is always beneficial. For instance, in industries with a significant environmental impact (e.g., chemicals and pharmaceuticals, tanning of textiles, and mining and refining of metals), the chief responsibility concerns frequently center around toxic waste disposal, water pollution, hazardous greenhouse
gas emissions, etc. In these cases, improving supply chain responsibility may involve both investing in cleaner equipment and working with the Tier 2 supplier to train its employees in the usage of clean technologies, implying joint efforts are required. Under these circumstances, our results show that supply chain responsibility can be improved by all forms of external stakeholder pressure.

By contrast, in many labor intensive industries, human rights violations are key sources of responsibility violations, e.g., forced or child labor or unsafe working conditions. The typical responsibility improvement activities in such cases, such as frequent supplier audits or financial subsidies to improve working conditions, are interchangeable and can be conducted by either the Tier 1 supplier or the manufacturer, implying substitutable efforts. Our results from the preceding sections shows that in such cases, when external stakeholders exert pressure with the intent of improving supply chain responsibility, their actions may backfire. Thus, we conclude that the nature of efforts—substitutable or complementary—plays a key role in generating the undesirable reaction of the manufacturer to increased external stakeholder pressure.

2.8 Conclusion

In supply chains with multiple tiers, downstream buyers are faced with a deceptively simple question: should higher tier suppliers be managed directly, or should their management be delegated to intermediate (e.g., Tier 1) suppliers? Delegation is less complex and, very often, more efficient from the point of view of total supply chain cost; conversely, direct control subverts the moral hazard problem inherent in delegation. In this paper, we have explored the consequences of this trade-off in the context of the management of social responsibility risk stemming from a Tier 2 supplier.

Our results have several implications. First, they provide guidance to downstream
firms on how to optimally choose an upper-tier supplier management strategy in the presence of responsibility risk. Under substitutable effort to reduce responsibility risk, the downstream manufacturer should pursue an all-or-nothing strategy: either delegate all management of Tier 2 to the Tier 1 supplier, or assume all management duties directly. Returning to the motivating examples discussed in the introduction, consider the incidents in the supply chains of General Motors and Mattel. In the case of GM, which experienced an explosion at the factory of a Tier 2 supplier, the effort required to avoid this disaster would include closer work with and monitoring of the Tier 2 supplier, which is likely less costly for Dicastal than GM, given the physical proximity of and familiarity between the companies; this suggests $k_1$ is much smaller than $k_0$. Moreover, the ultimate indirect loss caused by the explosion in Zhongrong’s factory to GM was likely small ($a_0$ is small), and Dicastal is a rather powerful manufacturer (it is the world’s largest wheel manufacturer, implying $r$ is large). These conditions combine to imply that delegation is likely optimal for GM. However, in the case of Mattel, which discovered lead paint in its products, the situation is quite different. The direct cost to Mattel stemming from the incident was substantial ($a_0$ is large), while the type of effort required to avoid the incident, i.e., testing products for hazardous materials, likely costs both Mattel and any of its Tier 1 suppliers a similar amount (i.e., $k_0 \approx k_1$). In addition, Mattel’s direct supplier (Lee Der Industrial Co.) is far less powerful than Dicastal, which suggests $r$ is small. These factors suggest that direct control of upper tier responsibility is likely optimal in the case of Mattel.

Second, we also show that, for the manufacturer, a penalty contract can coordinate the supply chain while simultaneously allowing the firm to extract all profit above the Tier 1 supplier’s reservation level; however, the optimal coordinating penalty contract might involve the manufacturer itself being penalized for responsibility violations. Importantly, penalty contracts allow delegation to achieve the
first best outcome when the Tier 1 supplier is more efficient at exerting effort than the manufacturer. Thus, we add to the literature on penalty contracts (Ang et al. (2014); Hwang et al. (2014); Plambeck and Taylor (2013)) by illustrating that such contracts are potentially extremely useful in the management of responsibility risk, given that global supply networks are increasingly long and complex, making direct control of upper tier responsibility management impractical for many manufacturers. However, as we discuss, there are a number of practical obstacles to the implementation of such contracts that may make them infeasible in many real world scenarios.

Third, as much of our analysis centers around the impact of external stakeholder pressure from consumers, NGOs, and policymakers, our results add to the growing supply chain literature that illustrates how these parties should (and should not) exert pressure to attempt to influence the responsibility level of a supply chain (Kalkanci et al. (2012); Plambeck and Taylor (2013); Guo et al. (2014); Xu et al. (2015)). Intuitively, in a centralized supply chain (or one coordinated via penalty contracts), increased pressure from any of these stakeholders always results in greater supply chain responsibility. In a decentralized supply chain under substitutable efforts, however, when the manufacturer optimally chooses between the delegation and control strategies, greater pressure from external stakeholders can lead to a reduction in supply chain responsibility. When effort is complementary, greater external pressure always leads to greater responsibility. This suggests that external stakeholders should pay close to attention to the relationships in a supply chain as well as the type of effort (complementary or substitutable) required to reduce the chance of violations. Awareness of this issue can maximize the effectiveness of external stakeholder pressure in improving supply chain responsibility.

Lastly, our results show how the multi-tier structure we have analyzed can lead to new insights in principal-agent settings (Yang et al. (2009); Babich and Tang
(2012); Kayis et al. (2013)). Because the effort is exerted in the facility of a third party (i.e., the Tier 2 supplier), while the manufacturer can act as a principal in inducing the agent (i.e., the Tier 1 supplier) to exert effort, the manufacturer can also bypass the Tier 1 supplier and exert responsibility improvement effort directly. The manufacturer’s optimal choice between these two approaches leads to new qualitative effects (such as external pressure reducing supply chain responsibility) that are not present in a supply chain with only two firms or when the manufacturer is forced into a single strategy. Hence, in addition to reinforcing the notion that either delegation or control may be optimal (Kayis et al. (2013)), our results illustrate that switching between strategies may lead to interesting and, potentially, problematic effects.

As supply chains continue to swell in both length and breadth, the management of multi-tier supply chains is an area that will undoubtedly only grow in importance in the coming years. Our model thus serves as both a first step in analyzing responsibility management in this setting, as well as an illustration of some of the important implications of managing suppliers in extended supply chains.
Supply Chain Planning for Random Demand Surges: Reactive Capacity and Safety Stock

3.1 Introduction

A central challenge in supply chain management is to develop system capabilities to match supply with demand in a cost-efficient manner. While there exist several demand models in the operations and supply chain literature that include random components, this paper considers a new and particularly challenging type of demand uncertainty – random demand surges. These are unpredictable yet substantial increases in demand in an otherwise relatively stable environment. Consider the following three illustrative examples:

- **Severe weather and disasters.** Severe weather (e.g., a hurricane or tsunami) and man-made disasters (e.g., oil spills, fires) generate unpredictable and large spikes in demand for both private firms as well as humanitarian organizations. Moreover, when disasters disrupt suppliers of critical industrial components, alternative suppliers of those components typically experience sudden and substantial demand surges. The combination of the rapid speed of globalization
and the increased frequency of natural disasters has exposed more firms to these types of demand surges.

- **Structural Demand.** Beyond normal demand derived from a firm’s own sales processes, firms occasionally secure sudden and large demands through external circumstances (e.g., new business partners, mergers, acquisitions). These structural events represent important and unique operational planning challenges due to their unpredictability and magnitude. For example, in our work with a high-tech Fortune 500 company, preparing for surges due to structural events was a far more pressing problem than managing normal demand fluctuations. Similarly, an executive at a large firm in the semi-conductor industry reported to us that one of their most difficult challenges is dealing with the unpredictability and infrequency of winning large contracts in developing countries that put enormous additional pressure on their supply chain.

- **New Product Introductions and “Viral” Products.** Finally, new product introductions typically create surges in demand for both retailers and suppliers. Especially for supply chains in the fashion and high-tech industries, the demand process can be well-conceptualized by a series of unpredictable surges associated with the release and/or sudden popularity of new products. One recent example is Apple’s release of the iPhone 6 and iPhone 6 Plus in 2014. Two months after the product launch, it was reported that only 58% of iPhone 6/6 Plus models were in stock at retail stores nationwide (Yarow (2014)). Further, social media has created the possibility for products to go “viral,” causing sudden and dramatic increases in demand (e.g., Shontell (2010)). For example, in December of 2012, a story on a small sweatshirt hoodie went “viral” on Twitter and Facebook, causing the small company American Giant to sell out in less than 36 hours (Manjoo (2013)).
A demand process with surges differs dramatically from a one with ordinary, stationary fluctuations. There are several new dimensions of uncertainty to consider: What is the likelihood of a surge demand occurring? What is the expected length and intensity of surges? What kind of trajectory will the surge demand follow? How difficult is it to predict surge demand characteristics? Our underlying demand model (see Section 3) captures these types of uncertainties. It is partly inspired by the state-space model of the environment (Song and Zipkin (1993)). However, a key difference is that our model provides an explicit characterization of the entire evolution of a demand surge, including its intensity, duration, and shape.

This paper examines a fundamental supply chain planning question that the particular nature of surge demands presents: How should an organization design its supply chain in order to position itself to efficiently cope with random demand surges? There are two fundamental ways a firm can prepare for surges: build reactive capacity or hold safety stock. Reactive capacity is the ability to ramp up production (above normal levels) when needed. To build reactive capacity, a firm can, for example, hold additional equipment and standby labor in order to be able to assemble product quickly if needed. Or, they can contract with an outside firm to give priority to their needs. On the other hand, safety stock is inventory that is held at all times that can be deployed when a surge occurs. To keep safety stock for demand surges, a firm can set aside extra inventory in their own warehouses. Or they can outsource safety stock by contracting with another vendor to hold the inventory for them. Of course, one can also employ a combination of reactive capacity and safety stock. For example, the high-tech retailer mentioned earlier not only holds extra machines in its service centers, but also compensates its vendors to hold additional components so that the vendors can quickly assemble the components into machines if needed. Similarly, to prepare for disasters, humanitarian and governmental organizations both maintain safety stocks of key relief items at their own warehouses and contract
with manufacturers for stand-by access to their production capacity should a disaster strike (e.g., see NASPO (2013)).

The main objective of the present paper is to better understand the optimal combination of reactive capacity and safety stock and how it depends on the type of surge demands the supply chain faces. To accomplish this goal, we construct an analytical model. There are two policies to determine: a planning policy (i.e., how much reactive capacity and safety stock to maintain), and a deployment policy (i.e., how to use the reactive capacity and safety stock to satisfy surge demands). We are able to characterize the optimal deployment policy and then obtain the optimal planning policy. The model is somewhat stylized, yet it includes the essential features of real systems. We capture reactive capacity via a fixed maximum production rate with a linear reservation cost. Safety stock is immediately available and carries a linear holding cost. In this way, the model captures two fundamental tradeoffs: On one hand, safety stock is immediately available while capacity reservation requires some ramp-up time; on the other hand, reactive capacity can generate product for a long time while only a limited amount of safety stock is available. We assume, plausibly, that surge demand must be met immediately or is lost forever. The objective is to minimize total cost, subject to a minimum service level constraint.

Our main contributions are as follows. First, the new demand model captures demand surges in a more nuanced manner than traditional models in the operations management literature. By specifying surges through a set (countable or uncountable) of sample paths, our model is remarkably flexible and can capture heretofore unstudied aspects of demand uncertainty, such as shape. These features are important due to the lost-sales character of surge demands. It also allows a more precise and explicit tradeoff comparison between safety stock and reactive capacity; the optimal way to prepare for some demand shapes may be significantly different than for others, even if demands are identical in terms of simpler measures such as size.
Second, we are able to characterize the optimal planning policy analytically. To do this, we first find the optimal deployment policy for any fixed reactive capacity and safety stock levels and characterize its cost. We are able to prove that the service level constraint (which in its original form requires calculating the probability that demand is satisfied at every point in time during the surge demand) is equivalent to another constraint that only requires evaluating the probability of satisfying demand at one point in time, which we refer to as the critical time. This simplification is key to making a seemingly hard problem tractable. Finally, we show that the optimal policy type is determined by two thresholds that define four regions – do nothing, reactive capacity only, safety stock only, or both – and we demonstrate how to evaluate the best policy in each region.

Third, we conduct extensive sensitivity analysis to better understand how different surge demand/product characteristics (and the ability to accurately predict them) impact the optimal planning policy. Specifically, we leverage several forms of stochastic comparison to illuminate how surge demand characteristics (i.e., duration, intensity, frequency, shape) affect the optimal absolute and relative levels of reactive capacity and safety stock. Table 3.1 in subsection 5.3 provides a high level summary of the results. Some results are intuitively appealing. For example, we find that one should favor reactive capacity as surge demands last longer or their durations become more uncertain. In contrast, one should invest more heavily in safety stock for short and intense demand surges. Other insights provide structural intuition. For example, we find that the two strategies remain proportionally attractive as the intensity of surge demand grows, so long as a particular type of stochastic equality holds. Finally, some insights are counter-intuitive. For example, we find that if an organization seeks to achieve a higher service level, while they should always increase reactive capacity, they may actually want to decrease safety stock. And, if safety stock is less expensive than reactive capacity, it is optimal to use safety stock
only when surge demands are very frequent or very infrequent, but not necessarily in between. This set of results offers clear “strategic guidelines” which managers can understand and use.

The remainder of the paper proceeds as follows. Section 2 reviews related literature. Section 3 presents the demand model and formulates the optimization problem. Section 4 derives the optimal policy. Section 5 contains sensitivity analysis and stochastic comparisons. Section 6 discusses extensions and makes some concluding remarks. All proofs are in the appendix.

3.2 Literature Review

The paper is related to those that study multiple supply modes to hedge against uncertainty. Several papers consider capacity reservation or backup agreements between a buyer and its supplier in the presence of supply risk or demand uncertainty. Such agreements can be designed to provide flexibility in industries where the demand is highly volatile and difficult to forecast. For example, Eppen and Iyer (1997) consider a backup agreement between a catalog company and its manufacturers and show that the backup option can increase the committed order quantities. Henig et al. (1997) study a backup contract between an automobile company and its supplier. Capacity reservations too can be designed to mitigate supply and disruption risks for the buyer. Literature in this stream includes Tomlin (2006), Chopra et al. (2007), Hou et al. (2010), and Saghafian and Van Oyen (2012). Our work differs in the way we model surge demands and in the focus on the strategic differences between the two supply modes: reactive capacity is slower but more flexible than safety stock.

Procurement strategy under supply uncertainty and disruption has received a great deal of attention recently. Based on the type of supply uncertainty addressed, most of this work can be differentiated into three groups: random yield ( Henig
and Gerchak (1990), Anupindi and Akella (1993), Federgruen and Yang (2009)), random capacity (Ciarallo et al. (1994), Erdem et al. (1999)), and random disruption (Parlar and Perry (1996), Tomlin (2006), Babich et al. (2007)). Our work is related to these papers in the sense that all consider procurement strategy under extreme uncertainty. However, while the literature deals with extreme uncertainty on the supply side, our work addresses extreme uncertainty on the demand side.

Our state-dependent demand process is related to the Markov-modulated demand process widely adopted in the inventory literature, in which an underlying Markov chain describing the demand environment drives the demand distribution. See, for example, Song and Zipkin (1993), Sethi and Cheng (1997), and Chen and Song (2001). Relevant to our paper, Kapuściński and Tayur (1998) and Aviv and Federgruen (1997) have investigated optimal inventory policies for cyclic demand when the supply process has finite capacity, and Bhat and Krishnamurthy (2014) consider the optimal control of a make-to-stock system with exponential processing time and a Markov-modulated Poisson demand process. This stream of research shows that the optimal inventory policy should be state-dependent. In our model, the underlying demand environment is a renewal process, switching between normal and surge states. While the transition from the normal state to the surge state is Markovian, the transition in the other direction is not. Also, we use a sample-path approach to model the non-Markovian demand evolution in the surge state. Nonetheless, we too follow a state-dependent policy but focus on the policy in the surge state (by normalizing the policy parameters to zero for the normal state). In addition, we consider a dual sourcing problem—we decide both inventory and capacity reservation levels. In this regard, our work is related to somewhat Song and Zipkin (2009), who study a multi-sourcing inventory problem with Poisson demand. While they focus on how to deploy the alternative sources in a Markovian world, we consider both high-level planning and lower-level deployment decisions in a non-Markovian environment.
Finally, we note that planning for surge demands is of particular interest for disaster preparedness. There has been a growing interest in studying some of the unique supply-chain problems that disasters create and how to apply or extend our existing theories to these challenges (see Van Wassenhove (2006) and Ergun et al. (2013)). Disaster supply chains are special in several ways including the supply process, the demand process, financial resources, the participating players, and the information technology available (Ergun et al. (2013)). Although our model is not specific to disaster preparedness, our new way of representing surge demands does address some aspects of disasters not covered in standard models.

3.3 Model

We begin by describing the surge demand process for a single product. Then, we formulate the joint safety stock and reactive capacity planning problem. Throughout the paper, \( y^+ = \max\{y, 0\} \), \( x \land y = \min\{x, y\} \), and \( 1_{\{\}} \) is the indicator function. In general, random variables are denoted with capital letters.

3.3.1 Demand Process

Consider an environment with a normal and a surge state. (We discuss the extension to more normal and surge states in Section 6.) The environment begins in normal state, during which there is a constant demand rate. We assume the normal demand rate is zero without loss of generality. The system stays in the normal state for a random period following an exponential distribution with rate \( \lambda \). Once it enters the surge state, the environment stays there for a random time \( T \). After \( T \), the environment returns to the normal state. We denote \( \tau = E[T] \) the expected duration of the surge state and \( \nu = 1/\lambda \) the expected time in the normal state.

A salient feature of our modeling framework is a novel way to describe surge demands. We assume the demand in the surge state is a continuous-time stochastic pro-
cess \{X(t), t \geq 0\}. Instead of assuming parametric distributions of \(T\) and \{\(X(t), t \geq 0\)\} (such as a lognormal distribution for \(T\) and a Brownian motion for \(X\)), we take a sample-path approach. Let \((\Omega, \mathcal{F}, \mathbb{P})\) denote the probability space for all possible demand sample paths (or scenarios) in the surge state. Each \(\omega\) corresponds to a unique forecasted demand sample path. More specifically, each \(\omega \in \Omega\) specifies surge duration \(T(\omega)\), the corresponding demand trajectory \{\(X(t, \omega), 0 \leq t \leq T(\omega)\)\}. Measure \(\mathbb{P}(\omega)\) gives the probabilities of these \(\omega\)s. The sample space \(\Omega\) can be either countable or uncountable.

This sample-path definition of surges is consistent with the observation that demand forecasting usually provides a variety of possible scenarios along with a probability of each scenario (Sheffi et al. (2005)). It is also complimentary to the growing trend of “big data” and increasingly-common forecasting methods through statistical tools, such as those that rely on the bootstrapping method. Indeed, many companies and government agencies maintain data bases containing long histories of trajectories of surge demands and the incidents that triggered them. They also have the analytical capability to generate probabilistic predictions for future demand surges based on this data (Sheffi et al. (2005)).

Figure 3.1: Depiction of possible surge demand paths
Depending on the triggering event and the surge product, the surge demand may have any of several possible patterns. See Figure 3.1 for an illustration of sample demand paths. Here, some surges start with a high intensity but do not last long, as in $x(t, \omega_1)$, while others last longer with gradually increasing intensity, as in $x(t, \omega_2)$. Surge instances may peak early, as in $x(t, \omega_3)$ or later, as in $x(t, \omega_4)$. This model is general and flexible enough to represent a wide variety of surge demand evolutions. The following examples suggest some of the possibilities:

. (Rectangular Surge Demand) A rectangular surge demand has a sample path of a rectangular shape. That is, it stays flat over the surge duration. Denote $\omega = (l, s) \geq 0$ where $l$ and $s$ represent the random length and magnitude of surge demand, then $X(t, \omega) = s, 0 \leq t \leq T(\omega) = l$. This shape may be used as a key element to estimate surge demand in response to a flood, which is often relatively stable over duration.

. (Triangular Surge Demand) A triangular demand has a sample path of a triangular shape, which is characterized by its duration $l$, peak time $\beta$ and demand magnitude $s$ at the peak time $\beta l$. Let $\omega = (l, \beta, s)$ where $l, \beta, s$ are positive random variables with $\beta \leq 1$. Then $T(\omega) = l$, and

$$ X(t, \omega) = \begin{cases} \frac{s l}{\beta} t, & 0 \leq t \leq \beta l \\ \frac{l-t}{(1-\beta)} s, & \beta l < t \leq l \end{cases} $$

This shape may be used as a key element to estimate surge demand triggered by an earthquake, which spikes sharply at some point.

Of course, we can easily define other shapes by extending the space of $\omega$. These simple demand evolution patterns can also be used as constituent elements for more complex demand processes. Nevertheless, the crucial benefit of this surge demand model is not that it can accommodate several of these parametric types of demand
surges. Rather, to foreshadow our analytical results in Section 5, the sample-path definition of surge demands facilitates stochastic comparisons via sample-path arguments.

For tractability, we assume that any realization of the demand trajectory \( \{X(t, \omega), 0 \leq t \leq T(\omega)\} \) is continuous and uni-modal with respect to \( t \). This assumption allows us to solve the optimal deployment policy. However, it is also reasonable: Even though the surge demand may manifest some small oscillations locally in reality, it usually maintains the unimodal property globally. Finally, we note that in the rest of the paper, when \( \omega \) is given, we often suppress \( \omega \) for notational simplicity. For example, we write \( T \) instead of \( T(\omega) \) and denote \( x(t) = X(t, \omega), 0 \leq t \leq T(\omega) \), when the context is clear.

3.3.2 Problem Formulation

We consider two operational levers: safety stock and reactive capacity. A safety stock of size \( m \) guarantees the immediate availability of up to \( m \) units in inventory when surge state occurs. Reactive capacity of level \( \mu \) guarantees the maximum available production rate \( \mu \) to produce the product in need while in the surge state. Let \( (m, \mu) \) denote this joint planning policy. Once in the surge state, the manager needs to make a production decision for how to use the reserved capacity at any time \( t \). We call the inventory produced from this production the reactive stock. Finally, the manager must decide how to use the safety stock and the reactive stock to meet surge demand. These decisions comprise the production policy from reactive capacity and deployment policy.

Because the production and deployment problem depends on the planning policy, we proceed by formulating the problem backwards: first formulating the production and deployment problem for a fixed \( (m, \mu) \) before formulating the planning problem.
Stage 2: Production and Deployment Problem

Under any given joint planning policy, at each time point \( t \) in the surge state, the decision-maker needs to decide a production policy \( p_\mu(t) \) for how to produce from the reserved capacity, and a deployment amount \( y_m(t) \) from the remaining safety stock and \( y_\mu(t) \) from the remaining reactive stock. Let \( (y_m(t), y_\mu(t) | 0 \leq t \leq T(\omega)) \) denote a deployment policy. To determine the optimal deployment policy, we need to specify the economic factors associated with different operational levers. Safety stock incurs a holding cost \( h \) per unit/time. Capacity reservation costs \( r \) per unit/time. We also assume linear purchasing costs: it costs \( c \) to purchase units from safety stock and \( w \) to purchase units produced with reactive capacity. We focus our analysis on the case \( c < w \) due to the fact that the marginal production cost for a supplier is typically greater when he needs to ramp-up production with short notice. Nevertheless, similar analysis can be repeated for \( c \geq w \) and most of our results still hold, although we omit those details here.

Given a planning policy \((m, \mu)\), let \( C(m, \mu) \) denote the optimal expected deployment costs. It can be expressed as follows:

\[
C(m, \mu) = E_\omega \left[ \min_{p_\mu(t), y_m(t), y_\mu(t) \in \mathbb{R}^+} \left\{ c \int_0^{T(\omega)} y_m(t) \, dt + w \int_0^{T(\omega)} y_\mu(t) \, dt \right\} \right] 
\]

s.t. \( \max P \left\{ \omega : \int_0^t X(s, \omega) \, ds \leq y_m(t) + y_\mu(t), \ 0 \leq t \leq T(\omega) \right\} \) (3.1)

\[
y_m(t) + y_\mu(t) \leq X(t, \omega) \quad (3.2)
\]

\[
y_m(t) = 0, \text{ if } \int_0^t y_m(s) \, ds = m \quad (3.3)
\]

\[
y_\mu(t) \leq p_\mu(t), \text{ if } \int_0^t y_\mu(s) \, ds = \int_0^t p_\mu(s) \, ds \quad (3.4)
\]

\[
p_\mu(t) \leq \mu, \ 0 \leq t \leq T(\omega)
\]

Constraint (3.2) guarantees that the supply does not exceed the demand. Constraint (3.3) specifies that the total deployment from safety stock cannot exceed the total
safety stock. Constraint (3.4) says that when the cumulative deployment from reactive stock by $t$, $\int_0^t y_{\mu}(s) \, ds$, reaches the maximal possible reactive stock $\int_0^t p_{\mu}(s) \, ds$, the planner can only deploy up to the current production capacity $p_{\mu}(t)$, which is limited by the reactive capacity level $\mu$. In the deployment stage, the planner aims to satisfy the demand as much as possible, which is captured by the service level maximization constraint (3.1). Here, we have defined the service level as the probability that all surge demand is immediately satisfied. Therefore, like the in-stock rate, a 95% service level is interpreted as there being a .95 probability that all surge demand is served upon arrival. On top of service level maximization, the planner desires to satisfy the demand in a cost-efficient way. Note that the production and deployment decisions at time $t$ are made when the demand trajectory after $t$ is not available.

**Stage 1: Planning Problem**

Now, given the costs from the production and deployment problem, the expected total cost rate function of the joint planning policy can be written as

$$G(m, \mu) = \frac{1}{\nu + \tau} \left( [hm + r\mu] \tau + C(m, \mu) \right) \quad (3.5)$$

It is the sum of two types of costs. The first term is an expected preparedness costs, which is the combined safety stock holding and reactive capacity reservation costs incurred during the normal state. Such an assumption is reasonable because the safety stock is being depleted and production capacity is being deployed during the short time in the surge state (relative to the entire cycle time) so that the holding and reservation cost is minimal. (It is straightforward to alter the analysis to account for the preparedness costs both during the normal and surge state and the main results still hold.) The second term is the expected optimal deployment costs incurred during the surge state, which is obtained from Stage 2 above.
The joint planning problem is to minimize the long-run average system cost, under the surge demand coverage or service level constraint:

\[
A : \min_{m, \mu} G(m, \mu) \\
\text{s.t. } P\left\{ \omega : \int_0^T X(s, \omega) \, ds \leq y_m^*(t) + y_\mu^*(t), \ 0 \leq t \leq T(\omega) \right\} \geq \alpha \quad (3.6)
\]

The inequality (3.6) is the service level constraint with a service level lower bound \( \alpha \).

Another way to formulate this problem is to maximize surge demand coverage under a budget constraint of the long-run system cost \( \overline{G} \). We will later show that these two formulations are equivalent.

3.4 Optimization

The joint planning problem is difficult to solve in its original form \( A \); one must optimize over both the planning decisions \((m, \mu)\) and the deployment decisions \((y_m(t), y_\mu(t) \mid 0 \leq t \leq T(\omega))\) and determining the objective service level requires evaluating an inequality at every time point \( t \). Further, the optimal deployment decision is constrained by the underlying production policy \( p_\mu(t) \) that shapes the reactive capacity. We first address these challenges before presenting the optimal policy.

3.4.1 Optimal Deployment Policy

We begin by solving the problem backwards and characterizing the optimal production and deployment policy.

Lemma 11. If the total surge demand \( D(\omega) = \int_0^{T(\omega)} x(s) \, ds \) is known upon entering the surge state, then for a given planning policy \((m, \mu)\), the optimal production policy \( p_\mu^*(t) \) is to produce at rate \( \mu \) until either the cumulative stock produced plus the
The optimal deployment policy is greedy, always deploying from the cheaper source first. Specifically,

\[
(y^*_m(t), y^*_\mu(t)) = \begin{cases} 
(x(t), 0), & \text{if } m > \int_0^t y^*_m(s) \, ds \\
(0, x(t)), & \text{if } m = \int_0^t y^*_m(s) \, ds \text{ and } \int_0^t y^*_\mu(s) \, ds > \int_0^t y^*_\mu(s) \, ds \\
(0, p^*_\mu(t)), & \text{otherwise}
\end{cases}
\]

(3.7)

The service level is simplified as \( P \left\{ \omega : \int_0^t X(s, \omega) \, ds \leq m + \mu t, 0 \leq t \leq T(\omega) \right\} \).

For a depiction of the optimal production and deployment policy, refer to Figure 3.2. We assume that although the planner can not predict the evolution of surge demand \( x(t) \), he does have an accurate estimate of the total surge quantity, i.e. \( D = \int_0^{T(\omega)} x(s) \, ds \), i.e., the area below the curve. Assume \( D > m \). At each time point \( t \) on the horizontal axis, the planner only knows the curve before this point, i.e., \( x(s), s \leq t \). Once entering the surge state at \( t = 0 \), the optimal deployment policy purchases and deploys the stockpiled inventory first because it is less expensive.

Therefore, up to time \( T_0 \) where \( m = \int_0^{T_0} x(s) \, ds \) (as marked in the shaded area), the deployment rate \( y^*_m(t) \) is the demand rate \( x(t) \). In the mean time, because \( D > m \),
the optimal production policy produces at full rate, targeting to meet all demand. Thus at time $T_0$, up to $a = \mu T_0$ amount of reactive stock is available, which is the area of the shaded triangular. However, at time $T_0$, the reactive stock $a$ is still below the target reactive stock $D - m$, as illustrated in the figure, where we moved the shaded triangular area to a subarea of equal size (although of different shape) in the unshaded area. Thus the system still produces at maximal speed until time $T_1$, at which shortage begins to occur (as illustrated in the area marked $d$) and the planner switches to produce as much as possible to satisfy the current demand rate. Even though the capacity exceeds demand rate after time $T_2$, the system only produces the demand rate $x(t)$ since demand not immediately satisfied is lost. The total lost demand is $d$.

An important feature of the optimal deployment policy is that one only needs to know the total demand in order to execute it. We will make this additional assumption throughout the remainder of the paper. This feature allows us to maximize the service level without over-production and tractably characterize $C(m, \mu)$ in order to derive results on the optimal planning policy. Moreover, it is largely consistent with how surge demands operate in practice. For example, once a disaster strikes, a new contract is won, or a new product is launched, the associated organization can assess the surge demand to make an estimate of its volume. Of course, if one only has an erroneous estimate of total demand, then their production policy may be suboptimal. However, unless the initial demand estimate is extremely poor, such errors are unlikely to cause major cost differences as demand updates will typically provide enough accuracy before to avoid production starting or stopping sub-optimally. Furthermore, even if the planner overestimates the surge quantity to such an extent that the firm overproduces, the left-over reactive stock from over production may be returned or used to satisfy other demand.

58
Service Level Transformation. Next, we simplify the service level formulation. In problem $A$, it is difficult to optimize over the constraint (3.6) because one needs to check whether it holds at every moment in time over the random surge duration $T$. The following definition helps us simplify this constraint to make the problem tractable.

For any given $\omega$, we say a time point $T'$ is a critical time if given that inventory position at this point is positive then it is positive at all other points; namely surge demand is completely satisfied. Mathematically, $T'$ is the time point such that if $
int_0^{T'} x(s) \, ds \leq m + \mu T'$, then $
int_0^{T} x(s) \, ds \leq m + \mu t$, $0 \leq t \leq T(\omega)$.

Recall that the surge demand is unimodal, so its trajectory must be of one of the following patterns: 1) stationary, 2) increasing, 3) decreasing, or 4) first increasing and then decreasing. This observation, combined with the nature of the optimal deployment policy above, leads to the following result.

1. (Critical Time) For any given $\omega$, there exists a unique critical point $T'$, which is a function of the reactive capacity $\mu$. In particular, if $x(t)$ is stationary or increasing, then $T'(\mu) = T$. If $x(t)$ is first increasing then decreasing or decreasing, let $t_p = \arg\max\{x(t) : 0 \leq t \leq T\}$, then

$$T'(\mu) = \begin{cases} T, & \text{if } x(T) \geq \mu \\ \arg\{x(t) = \mu, \ t_p \leq t \leq T\}, & \text{if } x(T) < \mu < x(t_p) \\ t_p, & \text{if } x(t_p) \leq \mu \end{cases}$$

Figure 3.3 provides an illustration of the critical time. The notion of the critical time enables us to simplify the service level characterization as follows.

Lemma 12. The service level (5) is equivalent to

$$P\left\{\omega : \nint_0^{T'(\mu,\omega)} X(s, \omega) \, ds \leq m + \mu T'(\mu, \omega)\right\}.$$
This equivalent expression of the service level is much easier to evaluate than the original one, because for each $\omega$, we only need to check one inequality instead of checking the inequality for every moment in time. For instance, for rectangular surge demands, $T'(\mu, \omega) = l$. Thus, according to Lemma 12, we can rewrite the service level as $P\{\omega = (l, s) : sl \leq m + \mu l\}$. Namely, the service level can be determined by comparing the cumulative demand with the cumulative supply at the end of the surge. Similarly, for triangular surge demands, the critical time is $T'(\mu, \omega) = (\beta l + (1 - \beta) l \frac{s - \mu l}{s} 1_{[\mu < s]}).$ Thus, according to Lemma 12, we can rewrite the service level as $P\{\omega = (l, s) : \frac{sl}{2} - \frac{\mu^2 l}{2s} (1 - \beta) \leq m + \mu l\}$ if $\mu \leq s$ and 1 otherwise. We will return to these special cases in our analysis of the impact of surge demand characteristics in Section 3.5.

Cost Function Transformation  By leveraging the optimal deployment policy and the definition of the critical time, $C(m, \mu)$ can be simplified to

$$C(m, \mu) = E\left[w \left( (D - m)^+ - \left( B'(\mu) - m \right)^+ \right) + c[D \wedge m] \right] \quad (3.8)$$
where $D(t) = \int_0^t X(s, \omega) \, ds$, $0 \leq t \leq T(\omega)$ and $B'(\mu) = D(T' (\mu)) - \mu T' (\mu)$. Consequently, problem $A$ can be simplified to

$$A': \min_{m, \mu \in \mathbb{R}^+} G(m, \mu)$$

s.t. $P \left\{ \omega : \int_0^{T'(\mu,\omega)} X(s, \omega) \, ds \leq m + \mu T'(\mu,\omega) \right\} \geq \alpha \quad (3.9)$

It can be shown that at optimality, the constraint (3.9) is tight, from which we obtain an analytical relationship between $m$ and $\mu$. Thus, $A'$ can be reduced to a univariate optimization problem. However, it is still intractable because the convexity of the objective function is not guaranteed. To overcome this difficulty, we introduce a unit penalty cost $b$ for each unsatisfied demand $[D(T') - \mu T' - m]^+$. In this way, we can transform the cost $C(m, \mu)$ and obtain the associated transformed total average cost is

$$G^p(m, \mu) = \frac{1}{\nu + \tau} ((hm + r\mu) \tau + C^p(m, \mu)) \quad (3.10)$$

$$C^p(m, \mu) = cE[D] + (w - c) E[D - m]^+ + (b - w) E[B'(\mu) - m]^+ \quad (3.11)$$

The corresponding transformed optimization problem is

$$B : \min_{m, \mu \geq 0} G^p(m, \mu) \quad (3.12)$$

which we refer to as the penalty cost formulation.

### 3.4.3 Optimal Planning Policy

We proceed by first solving the penalty cost formulation in problem $B$. Later, we prove that solving problem $B$ is equivalent to the original service level constraint formulation $A$. The function $G^p(m, \mu)$ is well-behaved. Specifically,
Lemma 13. \( G^p(m, \mu) \) is jointly convex in \((m, \mu)\). The optimal solution \((m^*, \mu^*)\) to the problem \( B \) exists and can be obtained by solving the Kuhn-Tucker first-order conditions.

We can show that the optimal joint planning policy is one of the following four types depending on the parameter range: (i) no action; (ii) employ safety stock only; (iii) employ reactive capacity reservation only; (iv) employ a joint strategy. (The optimal joint planning policy may not be unique, but only under some trivial cases, which we ignore.) To determine these regions and evaluate the optimal policy, we first make the following definitions.

When \( h < \frac{b-c}{\nu} \), define \( m_1(h) \) as the solution to

\[
P \{ D > m_1 \} = \frac{h \nu}{b - c},
\]

(3.13)\)

When \( r < \frac{(b-w)\nu}{\nu} \), define \( \mu_1(r) \) as the solution to

\[
E \left[ T' \left( \mu_1 \right) \left| B' \left( \mu_1 \right) > 0 \right. \right] P \left\{ B' \left( \mu_1 \right) > 0 \right\} = \frac{r \nu}{b - w},
\]

(3.14)\)

define \((m_2, \mu_2)\) as the solution to

\[
(b - w) P \left\{ B' \left( \mu_2 \right) > m_2 \right\} + (w - c) P \{ D > m_2 \} = h \nu
\]

\[
E \left[ T' \left( \mu_2 \right) \left| B' \left( \mu_2 \right) > m_2 \right. \right] P \left\{ B' \left( \mu_2 \right) > m_2 \right\} = \frac{r \nu}{b - w}
\]

(3.15)\)

The following proposition provides a full characterization of the optimal planning policy:

Proposition 14. Let

\[
h^* \left( r \right) = \begin{cases} \frac{w-c}{\nu} + \frac{b-w}{\nu} P \left\{ B' \left( \mu_1 \left( r \right) \right) > 0 \right\}, & r < \frac{(b-w)\nu}{\nu} \\ \frac{b-c}{\nu}, & \text{otherwise} \end{cases}
\]

\[
r^* \left( h \right) = \begin{cases} \frac{b-w}{\nu} E \left[ T \left| D > m_1 \left( h \right) \right. \right] P \left\{ D > m_1 \left( h \right) \right\}, & h < \frac{b-c}{\nu} \\ \frac{(b-w)\nu}{\nu}, & \text{otherwise.} \end{cases}
\]
(i) If $r \geq r^*(h)$ and $h \geq h^*(r)$, then no-action is optimal, i.e., $(m^*, \mu^*) = (0, 0)$; 
(ii) If $r \geq r^*(h)$ and $h < h^*(r)$, then it is optimal to deploy safety stock only, i.e., $(m^*, \mu^*) = (m_1, 0)$;
(iii) If $r < r^*(h)$ and $h \geq h^*(r)$, then it is optimal to deploy reactive capacity only, i.e., $(m^*, \mu^*) = (0, \mu_1)$;
(iv) If $r < r^*(h)$ and $h < h^*(r)$, then it is optimal to deploy a joint strategy, i.e., $(m^*, \mu^*) = (m_2, \mu_2)$.

Moreover, $h^*(r)$ is increasing in $r$ and $r^*(h)$ is increasing in $h$.

Thus, the optimal planning policy is characterized by two switching curves, as depicted in Figure 3.4. As shown in the figure, it is optimal to employ a single operational lever if and only if the unit cost of that lever is below a certain threshold which depends on the unit cost of the other lever. For example, it is optimal to stockpile inventory only if and only if the unit inventory holding cost $h$ is below a certain threshold $h^*(r)$, as shown in the dark gray area. This threshold $h^*(r)$ is increasing in unit capacity reservation cost $r$. Namely, when reactive capacity becomes more efficient with lower unit reservation cost, inventory stockpiling needs also to have lower unit inventory holding cost to get deployed. Moreover, no action is taken when neither strategy is individually efficient ($h \geq \frac{b-c}{\nu}$ and $r \geq \frac{(b-w)\tau}{\nu}$) and

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.4.png}
\caption{Optimal policy region under case $c < w$}
\end{figure}
joint strategy is optimal when both strategies are relatively efficient \((h < h^* (r) \text{ and } r < r^* (h))\).

The optimal policy can be determined by the following steps: (1) first solve \(m_1 (h)\) and \(\mu_1 (r)\) according to (3.13) and (3.14); (2) based on those values, calculate the switching curves \(r^* (h)\) and \(h^* (r)\). (3) with these two switching curves, compare \(r\) with \(r^* (h)\) and \(h\) with \(h^* (r)\) to determine which action is optimal and calculate the corresponding optimal policy \((m^*, \mu^*)\).

The computation of the optimal policy involves numerically solving nonlinear equations (3.13)-(3.15). For simple surge patterns such as the rectangular or triangular surge demands, the nonlinear equations are easy to obtain. As a result, solving the optimal policy \((m^*, \mu^*)\) is equivalent to finding the unique zero point. Moreover, the nonlinear equations are monotone. Therefore, well established algorithms such as binary search could be applied here. For more complex surge demand patterns, the computation is more complex, requiring generating numerical approximations for the nonlinear equations. On the other hand, any surge demand pattern can be reasonably well approximated by a convex combination of the simple surge demand patterns, the solution of the latter is much simpler to obtain and can be used as a near optimal policy.

**Equivalence of Problem Formulations**

Finally, we verify that solving the optimal policy for the penalty cost problem \(B\) is equivalent to solving the other problems.

**Proposition 15.** There exists a one-to-one increasing function \(b(\alpha)\) such that the optimal solution \((m^*, \mu^*)\) to \(B\) under penalty cost \(b(\alpha)\) achieves the same optimal solution under the service level constraint \(\alpha\) in problem \(A\). Similarly, there exists a one-to-one increasing function \(b(\mathcal{G})\) such that the optimal solution \((m^*, \mu^*)\) to \(B\) under penalty cost \(b(\alpha)\) achieves the same optimal solution under the budget \(\mathcal{G}\).
In other words, the three formulations of the problem - minimizing cost given a target service level, minimizing total cost given a shortage penalty cost, and maximizing service given a fixed budget - are equivalent.

3.5 Impact of Product and Demand Characteristics

We now conduct sensitivity analysis in order to better understand how the optimal planning policy depends on the product’s cost structure as well as surge demand characteristics. We interested in the effect of these factors on the absolute levels of the optimal reactive capacity and safety stock, as well as their relative levels. To study the relative levels, we define

\[ \gamma_m = \frac{m^*}{m^* + E[\mu T' + D(T) - D(T')]} \], \quad \gamma_\mu = 1 - \gamma_m \]

Note that \( E[D - B'(\mu^*)] \) is the expected maximal available inventory from reactive capacity and \( m^* + E[D - B'(\mu^*)] \) is the expected available total inventory to meet surge demand. Therefore, \( \gamma_m \) measures the proportion of expected available inventory one plans to procure from safety stock. As \( \gamma_m \) increases (\( \gamma_\mu \) decreases), we say that safety stock (reactive capacity) is increasingly attractive.

3.5.1 Product Characteristics

We first investigate the impact of the cost parameters on the optimal policy.

**Proposition 16.** \( \mu^* \) is continuous and increasing in \( h, c \) but decreasing in \( r, w \), while \( m^* \) is continuous and decreasing in \( h, c \) but increasing in \( r, w \). Although \( \mu^* \) is increasing in \( b \), \( m^* \) and \( \gamma_m \) may be increasing or decreasing in \( b \).

Parameters \( h, c, r, w \) contain all cost information about the surge product, cost of production and holding/reserving. Depending on the type of surge product, unit
holding cost $h$ can be much higher than the physical and financial holding cost. For example, food and medical supply usually possess short life cycle and must be discarded once they expire. As a result, the replacement fee for ex-piracy becomes another significant part of $h$. $r$ contains the margin investment and maintenance cost for extra capacity, therefore is high for products with costly capacity investment. $c$ represents the mass production cost with sufficient planning time and under complete market competition. $w$ represents the ramp-up production cost of manufacturer under short response period. Usually, extra cost $w - c$ is small for commodity with short production lead time and automated production process, high for products with complex production process such as high tech consumer electronics. Proposition 16 demonstrates that the optimal planning policy is affected by cost parameters in an intuitive way. Each optimal level increases as its associated costs decrease relative to the other strategy. Therefore safety stock is suitable for products which are easy to reserve but expensive to ramp up, while reactive capacity is preferable for products costly to hold but can be produced quickly when needed.

What is more interesting is that while $\mu^*$ is always increasing the penalty cost $b$, the optimal safety stock level $m^*$ may actually be decreasing in $b$. Put in another way, in light of Proposition 15, $m^*$ may also be decreasing in $\alpha$ and $\bar{G}$. That is, it may be optimal to decrease the amount of stockpile if facing a higher service level requirement or a more generous budget. For an example of this phenomenon, see Figure 3.5. Here, $X$ and $T \sim \text{Uniform}[0,1]$, $\lambda = 0.05$, $c = 0.5$, $w = 1.7$, $h = 0.1$, $r = 0.04$), and $\lambda$ is selected so that the proportion of surge period over normal period $\tau/\nu = 2.5\%$. This result may seem somewhat surprising: if the cost of not meeting surge demand increases, why would one ever want to decrease the safety stock? The intuition is as follows: as the penalty cost is high, the planner seeks to obtain a higher service level. Although it may be cheaper to hold safety stock, reactive capacity may be better at securing a higher service level in a manner that outweighs the cost advantage of
Figure 3.5: The impact of the shortage penalty on the optimal policy

safety stock. We will examine the differences between the two strategies more closely in the rest of this section.

3.5.2 Demand Characteristics

Recall that what makes our demand model unique is that it can capture a variety of dimensions of surge demand uncertainty. In order to study the impact of these surge demand characteristics on the optimal planning policy, we will employ several notions of stochastic orders defined below. Consider two random variables $Z$ and $Y$ (which represent some characteristic of demand surges). We shall comparing their corresponding optimal policy parameters, such as $m^*(Z)$ and $m^*(Y)$, if $Z$ and $Y$ are stochastically ordered in some way.

. (Usual Stochastic Order) $Z$ is less than $Y$ in usual stochastic order, denoted $Z \leq_{st} Y$, if and only if $P(Z > x) \leq P(Y > x)$ for any $x \in (-\infty, \infty)$.

. (Convex Order) $Z$ is less than $Y$ in convex order, denoted $Z \leq_{cx} Y$, if and only if $E[\phi(Z)] \leq E[\phi(Y)]$ for any convex function $\phi : \mathbb{R} \rightarrow \mathbb{R}$.

Note that if $Z \leq_{cx} Y$, then $E[Z] = E[Y]$. Thus, the convex order can only be used to compare the uncertainties of variables with the same mean. To compare the uncertainties of variables with different means while holding uncertainty constant in
a stochastic sense, we use the following two less frequently used stochastic orders in the Operations Management literature.

. (Lorenz Order) $Z$ is less than $Y$ in Lorenz order, denoted $Z \preceq_{\text{Lorenz}} Y$, if and only if $\frac{Z}{E[Z]} \preceq_{\text{cx}} \frac{Y}{E[Y]}$.

Roughly speaking, the Lorenz order is similar to comparing the coefficient of variances of two random variables. In fact, $Z =_{\text{Lorenz}} Y$ implies that $Z$ and $Y$ have equal coefficient of variances. We will also leverage the fact that if $Z \preceq_{\text{Lorenz}} Y$, then there exists $k \geq 1$ such that $Z \sim kY$.

The next order enables us to compare uncertainties between variables with different means by adding or subtracting a constant.

. (Dilation Order) $Z$ is less than $Y$ in dilation order, denoted $Z \preceq_{\text{dil}} Y$, if and only if $Z - E[Z] \preceq_{\text{cx}} Y - E[Y]$.

For more information on these orders, see Muller and Stoyan (2002).

**Duration**

How does the surge demand duration affect the optimal planning policy? The following proposition shows the impact of a stochastically longer surge demand, controlling for uncertainty in the sense of the Lorenz order.

**Proposition 17.** Let $T_1$, $T_2$ be two surge durations with $T_1 \preceq_{\text{Lorenz}} T_2$ such that $T_2 \sim kT_1$ for some $k \geq 1$. Let $X_1(t)$, $X_2(t)$ be their corresponding surge demands with $X_2(t, \omega) = X_1 \left( \frac{T_1(\omega)}{T_2(\omega)} t, \omega \right)$, $\forall \omega \in \Omega$. Then, safety stock is less attractive under $T_2$ than under $T_1$ i.e., $\gamma_m(T_2) \leq \gamma_m(T_1)$. Moreover, $\mu^*(T_2) \geq \mu^*(T_1)$ and $m^*(T_2) \leq km^*(T_1)$.

The intuition behind Proposition 17 is as follows. Because reactive capacity is available in the entire surge duration, the expected unit reactive capacity cost $tv/\tau$
is lower for longer surge demands, leading to a larger \( \mu^* \). A stochastically longer duration also implies a larger total demand \( D \), which may result in higher optimal safety stock level, although this effect is smaller and bounded by \( m^* (T_2) \leq km^* (T_1) \).

The problem structure bears a special relationship with the Lorenz order that makes the analysis tractable. However, we find similar results with numerical studies with other types of stochastically longer surge demands. Specifically, we considered the case when surge duration \( T \) increases under the dilation order. Under the dilation order, there exists \( k \geq 0 \) so that \( T_2 \sim T_1 + k \). We set \( X \) and \( T \) to be Uniform\([0, 1]\), and set \( T = k + aX \), \( 0 \leq k \leq 1 \), where \( k \) varies by increments of .1 and \( a \in \{0.1, 1, 5\} \).

We also vary \( \lambda \) to be \( \{\frac{1}{20}, \frac{1}{10}, \frac{1}{5}, 1\} \) so that \( p = \frac{\tau}{\nu} \) falls in [2.5%, 50%]. We consider the cost parameters \( c \in \{0.1, 0.5, 1, 5\} \), \( w \in \{0.1, 0.5, 1, 5\} \), \( h \in \{0.01, 0.1, 0.5\} \) and . For each set of cost parameters, there exists \( r_{\text{min}} (h) \) and \( r_{\text{max}} (h) \) such that a joint strategy is optimal if and only if \( r_{\text{min}} (h) \leq r \leq r_{\text{max}} (h) \). Because we want to focus on these cases, we chose \( r \) is according to set \( \{r_{\text{min}} (h) + \frac{1}{4}(r_{\text{max}} (h) - r_{\text{min}} (h)), r_{\text{min}} (h) + \frac{1}{2}(r_{\text{max}} (h) - r_{\text{min}} (h)), r_{\text{min}} (h) + \frac{3}{4}(r_{\text{max}} (h) - r_{\text{min}} (h))\} \). We conducted a similar set of numerical experiments with exponential, normal and log-normal demand distributions. All numerical results suggest that when \( T \) increases in dilation order, \( \gamma_m \) decreases, as shown in Figure 3.6 (left).

Thus, we conclude that when surge duration increases, controlling for uncertainty, reactive capacity is more attractive. However, this is not necessary the case when \( T \) increases without control of uncertainty. Figure 3.6 (right) shows a counter example. Here, \( T \sim \text{Uniform}[k, 1], k \in \{0, 0.2, 0.4\}, X \sim \text{Uniform}[0, 1], \lambda = 0.05 \), \( c = 0.3, w = 0.8, r = 0.04 \) and \( h = 0.06 \). In this example, as \( k \) increases, \( \mu^* \) may decrease and \( \gamma_m \) may increase, especially when the required service level is high (\( b \) is large). The reason is that although longer surges increases the attractiveness of reactive capacity, second-order stochastic differences in durations may undermine that effect, as we will show next.
To investigate second-order effects, we leverage the convex order.

Proposition 18. Let $X_1(t)$ and $X_2(t)$ be two rectangular or triangular surge demands, where $T_1$ and $T_2$ are the respective surge durations such that $T_1 \leq_{cx} T_2 = k(T_1 - E[T_1]) + E[T_2]$, $k > 1$. Safety stock is less attractive under $T_2$ than under $T_1$, i.e., $\gamma_m(T_2) \leq \gamma_m(T_1)$. Moreover, $\mu^*(T_2) \geq \mu^*(T_1)$.

Proposition 17 shows that reactive capacity becomes more favorable when surge duration is more uncertain. The intuition is that in this case the risk of overage and underage associated with safety stock increases dramatically compared to the risk associated with reactive capacity. Although we must assume a specific parametric form of surge demands to prove the result, we believe the result is likely to be more general. Some evidence along these lines is that if we define $\Delta = C(m_1, 0) - C(0, \mu_1)$ as the cost difference between a (forced) safety stock only strategy and a (forced) reactive capacity only strategy, then we can show that $\Delta$ is greater under $T_2$ than under $T_1$ if $T_2 \geq_{cx} T_1$.

If the durations are uniformly or normally distributed, then Proposition 18 is applicable. To further verify the robustness of our findings under convex order, we conduct numerical experiments for the log-normal distribution $T \sim \text{ln}N$ with both rectangular and triangular surge demands, using parameters similar to those outlined.
in subsection 3.5.2. These experiments yield consistent results: when \( T \) increases in the convex order, \( m^* \) decreases but \( \mu^* \) increases. Together with Proposition 18, these results suggest that reactive capacity is more advantageous when \( T \) is more variable in convex order.

**Intensity**

Next, we investigate the effect of surge demand intensity. We obtain the following first-order effect.

**Proposition 19.** Let \( X_1(t) \) and \( X_2(t) \) be surge demands such that \( X_1(t) \leq_{st} X_2(t) \) and \( X_1(t) =_{Lorenz} X_2(t) \), \( 0 \leq t \leq T \). Then, safety stock and reactive capacity are equally attractive under \( X_1(t) \) and \( X_2(t) \), \( \gamma_m(X_2) = \gamma_m(X_1) \). Specifically, there exists \( k \geq 1 \) such that the optimal hedging policy \((m^*(X_2), \mu^*(X_2)) = (km^*(X_1), k\mu^*(X_1))\).

Proposition 19 states that when the intensities of surge demands increase, keeping Lorenz order equality (thus, keeping a constant coefficient of variation), the optimal stockpile and capacity levels increase proportionally. It reveals a special relationship between the structure of the optimal policy and surge demand uncertainty. If \( X(t) \) increases in another fashion, e.g., by a constant, the change in the optimal policy may differ from Proposition 19. The underlying logic is that the extra \( k \)th unit of intensity can be satisfied by reactive capacity without causing extra uncertainty. However, this extra \( k \)th unit of demand will incur extra \( kT \) demand uncertainty for safety stock. Thus, which strategy to favor depends on the tradeoff between unit cost differences and this uncertainty difference. When the cost difference \( w - c \) is small or when the required service level \( \alpha \) is low, the unit cost difference becomes less important and reactive capacity’s attractiveness increases. Otherwise, safety stock’s attractiveness increases. Numerical experiments (omitted here) confirm the
same trends under uniform, normal, exponential and log-normal distributions with rectangular and triangular surge demands. If intensity uncertainty grows without changing the expected intensity, numerical experiments (not reported here) indicate no clear trend for strategy preference.

**Shape**

In the previous two sections, we isolated duration and intensity to identify their impact on the optimal strategy. However, in isolation, these factors do not control for total demand. To study the impact of the surge demand shape, we now conduct stochastic comparisons while fixing total demand. Specifically, we investigate the impact of compactness, peak location and variability over duration.

**Compactness** First, we address the question: how does the optimal policy change when the surge demand is more or less “compact”?

**Proposition 20.** Let $T_1$ and $T_2$ be surge durations such that $T_2 \sim kT_1$, and define $X_2(t) = X_1\left(\frac{t}{k}\right)$, $0 \leq t \leq T_2$. Then, $\gamma_m(X_1, T_1) \geq \gamma_m(X_2, T_2)$. Specifically, $m^*(X_2, T_2) \leq m^*(X_1, T_1)$, $k\mu^*(X_2, T_2) \geq \mu^*(X_1, T_1)$.

Proposition 20 suggests that reactive capacity is preferred when the surge demand is of low intensity but lasts a long time. On the other hand, quick but intense surges favor the safety stock strategy. This result is consistent with the intuition that reactive capacity is more suitable for surge demands that allows it time to ramp-up, while safety stock is more valuable for immediate response.

**Peak Location** Next, we consider the effect of peak demand location. Recall that demand process $\{X(t, \omega), 0 \leq t \leq T(\omega)\}$ is continuous and unimodal. Assume $X(t, \omega)$ is increasing in time $t$ for each given $\omega$. Construct the following demand
group

\[ X'(t, \beta, \omega) = \begin{cases} 
X \left( \frac{t}{\beta}, \omega \right), & 0 \leq t \leq \beta T(\omega) \\
X \left( T - \frac{t-\beta T}{1-\beta}, \omega \right), & \beta T(\omega) < t \leq T(\omega)
\end{cases} \]

Here \( \beta \) captures the peak location: when \( \beta = 0 \), \( X'(t, \beta, \omega) = X(T-t, \omega) \), the demand is decreasing and peaks at 0; when \( \beta = 1 \), \( X'(t, \beta, \omega) = X(t, \omega) \), the demand is increasing and peaks at \( T \). This construction keeps the majority of demand sample path information while maintaining the fixed demand property under peak location \( \beta \) change. In particular, the total demand \( D(\beta, \omega) \) is constant with respect to \( \beta \) for each given \( \omega \). We have:

**Proposition 21.** Let \( \beta_1 \) and \( \beta_2 \) be two peak times for surge demand \( \{ X'(t, \beta, \omega), 0 \leq t \leq T(\omega) \} \) such that \( \beta_1 \leq \beta_2 \). Safety stock is more attractive under earlier peak times, \( m^*(\beta_1) \geq m^*(\beta_2) \) and \( \gamma_m(\beta_2) \leq \gamma_m(\beta_1) \).

Note that given fixed total demand, if \( m^* \) decreases while the service level \( \alpha^* \) increases, then \( \mu^* \) must increase. Thus Proposition 21 suggests that when surge demand peaks later, reactive capacity is more favorable. Interestingly, our numerical experiments show that it is possible for the optimal reactive capacity level to decrease if the unit penalty cost \( b \) is not large enough. This is because when the peak location is later, the demand ramps up slower, which makes the demand easier to satisfy from reactive capacity. If the service level \( \alpha^* \) is low, peaking later gives reactive capacity a chance to save by reducing the ordering quantity while still maintaining a larger service level.

**Stability/Volatility Over Duration** Third, we consider the stability of surge demand by maintaining the same total demand while adjusting the demand volatility over its duration. To capture stability in a unimodal setting, we assume the surge demand is a linear combination of rectangular and triangular surge demands. Consider
a triangular surge demand \( \{X(t, \omega), 0 \leq t \leq T(\omega)\} \) with zero end points \( X(0, \omega) = X(T(\omega), \omega) = 0 \), and define \( X'(t, \rho, \omega) = X_1(t, \rho, \omega) + X_2(t, \rho, \omega) \), where \( T_1(\omega) = T_2(\omega) = T(\omega) \), \( X_1(t, \rho, \omega) \) is rectangular surge demand and \( X_2(t, \rho, \omega) \) is triangular surge demand with \( X_1(t, \rho, \omega) = \frac{\rho}{2} X(t_p(\omega), \omega) \), \( X_2(t, \rho, \omega) = (1 - \rho) X(t, \omega) \).

Thus, \( 0 \leq \rho \leq 1 \) captures the stability of demand over duration. As \( \rho \) increases, the total surge demand \( D(\rho, \omega) \) remains constant, but the severity of the peak decreases. In the extremes, this yields a rectangular surge demand with \( \rho = 1 \) and triangular surge demand with \( \rho = 0 \).

**Proposition 22.** Let \( \rho_1 \) and \( \rho_2 \) be two stability parameters for \( X'(t, \rho, \omega) \) such that \( \rho_1 \leq \rho_2 \). Safety stock is less attractive when the demand is more stable \( m^*(\rho_1) \geq m^*(\rho_2) \) and \( \gamma_m(\rho_2) \leq \gamma_m(\rho_1) \).

This result implies that one should favor reactive capacity when surge demand is more stable, but favor safety stock when surge demand is more variable. The intuition is that safety stock provides more flexibility to meet sudden increases in demand, while reactive capacity increases in cost efficiency when demand is stable.

**Frequency**

Finally, we investigate the impact of surge frequency. For tractability, we focus on rectangular surge demands to derive analytical results. Define the ratio

\[
\kappa = \frac{r(b - c)}{h(b - w)\tau} = \frac{r(v + \tau)}{(b - w)\tau} \cdot \frac{h(v + \tau)}{b - c},
\]

which captures the cost difference between safety stock and reactive capacity. Note that \( \kappa \) is increasing in \( r \) and \( w \), but decreasing in \( h \) and \( c \). When \( \kappa > 1 \), reactive capacity has higher overage/underage cost ratio and hence is, in a sense, more expensive than safety stock.
Proposition 23. (i) If $\kappa \leq 1$, then $\gamma_m = 0$ for small $\lambda$ and $\gamma_m = 1$ for large $\lambda$. (ii) If $\kappa > 1$, then $\gamma_m = 1$ for $\lambda$ large or small, and $\gamma_m < 1$ otherwise.

Proposition 23 states that if the surge is either rare enough or frequent enough, a single hedging strategy is optimal. The intuition is as follows: the surge frequency primarily impacts the ratio of time spent between the surge and normal state. Thus, normal state cost difference (i.e., the difference between inventory holding and reactive capacity costs) will dominate when $\lambda$ is small, but surge state cost difference (i.e., the difference between purchasing costs and shortage costs) will dominate when $\lambda$ is large. Thus if $\kappa \leq 1$, when $\lambda$ small, normal state cost difference dominates and safety stock is not used ($\gamma_m = 0$). On the other hand, when $\lambda$ is big, surge-state cost difference dominates and reactive capacity is not used ($\gamma_m = 1$).

To further understand the impact of surge frequency and to complement Proposition 23, we conducted numerical experiments under four different types of distribution of $X$ and $T$. We assumed $X \sim Uniform[0, 1]$, $T \sim a \cdot Uniform[0, 1]$ and the same parameters for $c, w, r, b$ and $a$ as in subsection 3.5.2. For surge frequency $\lambda$, we similarly choose $p = \lambda/(\lambda+1)$ as in subsection 3.5.2. Finally, to control the impact of $\kappa$, we specifically choose parameter $h$ such that $\kappa \in \{0.2, 0.4, 0.6, 0.8, 1, 2, 5, 10, 50\}$. We find that when $\kappa < 1$, $\gamma_m$ is increasing in $\lambda$. This is because as $\lambda$ increases, the safety stock’s cost cost advantage in the surge state (because $c < w$) increases while its cost disadvantage in normal state (because $\kappa < 1$) decreases. On the other hand, we find that when $\kappa \geq 1$, $\gamma_m$ is typically first decreasing then increasing in $\lambda$. This non-monotonicity may seem counter-intuitive. The key observation is that when surges are extremely rare or extremely common, the differences in cost parameters will tend to determine the strategy. When surges are extremely rare, the differences in $h$ and $r$ (which favor safety stock when $\kappa > 1$) will be the determining factor. When surges are extremely common, the differences in $c$ and $w$ (which also tends
to favor safety stock when $\kappa > 1$) will be the determining factor. For intermediate surge demand rates, however, safety stock’s cost advantage in the surge state (because $c < w$) increases while its cost advantage in normal state (because $\kappa \geq 1$) decreases. The trend of $\gamma_m$ is determined by which cost difference is more dominant.

3.5.3 Sensitivity Results Summary

In this section, we have analyzed the impact of various product and demand characteristics on the optimal policy. Together, these results provide easy-to-understand guidelines for managers to support strategic decisions. Table 3.1 provides a high-level summary of the results obtained in the last two sections, although we note that the nuances of some results are lost and other results derived in the section cannot be fully represented in the table. Moreover, we note that the effects summarized here need not be mutually exclusive. For example, one can simultaneously face stochastically longer as well as less frequent surge demands. Our model can be used to investigate the combination of multiple factors simultaneously on a case by case basis.

3.6 Concluding Remarks

By developing a new demand model that captures demand surges and formulating an optimization problem with two important yet fundamentally different strategies to prepare for them, this paper provides an analytical framework to guide supply chain planning and design. The sample-path feature of the demand model also lends itself to shed light on how companies can leverage big data for coping with demand surges. We derived two switching curves that determine the optimal combination of reactive capacity and safety stock, and show how to evaluate them through a system of equations. Our sensitivity analysis shows the impact of the cost profile of each
Table 3.1: Summary of product and surge demand characteristics impact on the optimal planning policy. * indicates maintaining Lorenz order equality.

<table>
<thead>
<tr>
<th>Product/Surge Demand Characteristic</th>
<th>Impact on Optimal Planning Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SS Level</td>
</tr>
<tr>
<td><strong>Cost/ Objective</strong></td>
<td></td>
</tr>
<tr>
<td>Larger safety stock holding/purchasing costs</td>
<td>-</td>
</tr>
<tr>
<td>Larger reactive capacity reservation/purchasing costs</td>
<td>+</td>
</tr>
<tr>
<td>Larger budget/service level/shortage penalty</td>
<td>Depends</td>
</tr>
<tr>
<td><strong>Duration</strong></td>
<td></td>
</tr>
<tr>
<td>Stochastically longer*</td>
<td>Depends</td>
</tr>
<tr>
<td>More uncertain duration</td>
<td>Depends</td>
</tr>
<tr>
<td><strong>Intensity</strong></td>
<td></td>
</tr>
<tr>
<td>Stochastically more intense*</td>
<td>+</td>
</tr>
<tr>
<td>More uncertain intensity</td>
<td>Depends</td>
</tr>
<tr>
<td><strong>Shape</strong></td>
<td></td>
</tr>
<tr>
<td>Stochastically more compact</td>
<td>+</td>
</tr>
<tr>
<td>Stochastically earlier peak</td>
<td>+</td>
</tr>
<tr>
<td>Stochastically more stable</td>
<td>-</td>
</tr>
<tr>
<td><strong>Frequency</strong></td>
<td></td>
</tr>
<tr>
<td>More frequent surges</td>
<td>Depends</td>
</tr>
</tbody>
</table>

strategy as well as several first and second order characteristics of surge demands on the optimal joint strategy. Together, these results provide a set of insights and managerial guidelines for demand surge planning across different industries.

We conclude with some opportunities for future research. We note that the focus our paper was to capture the key essence of the planning problem in order to gain high-level insight into the fundamental trade-offs between the two hedging strategies. As such, some assumptions served as simplifications or approximations of what would occur in practice. In some ways, the model can extended without much further complication. For example, we can extend the model to incorporate multiple normal and multiple surge states under certain assumptions. Multiple normal states could capture different levels of risks of entering different types of surges, each with its own set of possible sample paths. If the planner is well informed about the system
state and demand surge is rare enough to allow planning policy adjustments in every normal state, than a state dependent planning policy can be solved in the same way as in the main model by extending the set \( \Omega \). At the other extreme, if the planing policy cannot be changed over many demand surges, then a long-run state-independent planning strategy can also be solved in the same way using steady state analysis. Other extensions to the model may require significant further analysis. For example, the details of the optimal deployment policy during a surge demand are simplified in our model and one could investigate a more sophisticated dynamic program that simultaneously updates demand forecasts to make decisions. Including these details in the model would distract from the focus of the present paper, which is to characterize the basic properties of the optimal planning policy, but would be interesting to study in order to help practitioners execute deployment during a surge demand. Similarly, one could add more features and caveats associated with reactive capacity and safety stock (e.g., fixed lead times, ramp-up times for safety stock, etc.). Such analysis is useful in that it would aid practitioners determine the exact reactive capacity and safety stock levels, but likely at the cost of being able to derive high-level strategic insights such as those gained through our stochastic comparisons.
4.1 Introduction

As companies tap into the global market, the links between companies and its customers become more apparent. Any disfunction from the customers may cause devastating consequences for the firm’s operation. A simple example is the financial crisis in 2008 during which the economies in North America and Europe were severely hurt. Consequently, the purchasing power in these countries dropped substantially. As a result, thousands of Chinese factories in Guangzhou and Shenzhen, which produce consumer and electrical products such as apparel, toys, and small electronic appliances and export them to these countries, suffered significantly in sales and profits. Many of these factories were not able to maintain cash flow balance and quickly went bankrupt.

According to a survey (Economist Intelligence Unit, Survey (2013)), more than 60% of the company executives (C-level) voted that a paramount goal of theirs in recent years is to build up a resilient supply chain that can deal with the customer
risks and hedging against the customer risks has become as important as, if not more than, hedging against supplier risks.

There are generally three types of customer risks that a supplier needs to address. The first type is the disruption risk. A disruption at the customers can happen for many reasons. It can be any natural disaster, such as an earthquake or a fire, which ruins the customers’ production facility. Or it can be a man-made event, such as a labor strike in the customers’ factory, a shortage of other critical components for producing the customers’ products. The disruption risks are exogenous. Once they occur, the customers would temporarily cease ordering from the supplier. The global financial crisis in 2008 can be seen as a disruption for the Chinese suppliers. The second type of customer risks is the retention risk, which means a customer permanently leaves the supplier. For example, the customer may choose to switch to other suppliers if they are not satisfied with the service level they receive (Fitzsimons (2000)). Or it may be the customers’ operational strategy to backward integrate (with alternative suppliers) or discontinue their own production. This risk is endogenous in the sense that the probability of a customer’s leaving is dependent on the service level the supplier provided. The better the service level, the more profitable a customer can be, and therefore the more likely the customer will to stay. The third type of customer risks is the usual, but can be substantial, uncertainty in the customers’ order variability. For example, from our interactions with a Fortune 500 tech company, we observe that the demand in the electronics market manifests significant uncertainty and seasonality, especially when the customers launch new products.

To hedge against these risks, companies need to tradeoff two types of strategies: *acquisition* and *retention*. Many companies choose to rely heavily on acquisition to attract more customers and expand the client pool. According to Marketing Budget Survey (Survey (2009)), 56% of firms’ marketing budget have been spent on acquiring more customers, through advertising, promotion, etc. The underlying idea
of the acquisition strategy is the same as the multiple sourcing strategy to hedge against supply risks: with a larger customer pool, when some customers order less or get disrupted, others may fill the gap. While acquisition lowers a supplier’s risk of capacity under utilization and decreased revenue, it splits up the supplier’s limited resources and capacity when the total demand exceeds the supply, resulting in lower customer service levels and hence higher retention risk. Thus, how to balance between the acquisition strategy to diversify risks and the retention strategy to hedge against the retention risk is a non-trivial question.

Indeed, while the majority of the companies (71%) focus on the acquisition strategy, some companies choose to cultivate and retain the existing customers (Econsultancy, 2014).

Adding to this complexity is the fact that companies typically have several types of customers, with different risk profiles (demand uncertainty, loyalty or retention risk, disruption risk), order volumes, profit margins, acquisition costs, etc. Higher volume customers can usually demand lower prices due to both the scale economy effect and higher bargaining power. Thus, the optimal customer base selection dose not only depend on the number of customers, but also on the customers type. Moreover, when capacity shortages occur, companies need to make the operational level decision of capacity allocation. There is short-term profit gain by prioritizing the smaller but more profitable customers. However, this might damage the relationship with less profitable but bigger customers and risk losing their future business.

To better understand the tradeoff, the present paper aims to answer the following questions:

1. How should companies optimally allocate capacity among customers when capacity shortages occur, given that this policy affects customer retention risks? Which types of customers and orders should be prioritized or de-prioritized?
2. What should be the optimal customer base when taking into account various types of customer risks? In other words, what is the optimal balance between the acquisition and retention strategies?

3. How does market and customer characteristics affect the optimal polices?

We answer these questions by analyzing a supply chain model of a single supplier (the firm) with finite capacity selling to multiple types of business customers over time. The customer types differ in terms of profit margin, demand volume, risk profile, etc. Customers of the same type share the same profit margin and demand distribution. The firm faces three types of risks from its customers as discussed above. The customer disruption risk in each period follows a Bernoulli distribution, with an exogenous, customer-type dependent disruption probability. The customer retention risk in each time period also follows a Bernoulli distribution, but the probability that the customers stays with the firm in the current period is endogenous, which is an increasing function of the service level they received in the last period. In addition to these risks, the demands (or the orders) from each type of customers follow a common stationary distribution. At the beginning of each period, the supplier can choose to acquire more customers at a cost, which is constant for each given customer type. The supplier’s objective is to maximize expected profit over a planning horizon.

We first consider the problem with an infinite horizon. This setting represents a mature and well established market where the products are commodity type with long life cycle. We first show that the optimal capacity allocation policy is stationary, and we derive an explicit optimal allocation solution. Consistent with the literature (Adelman and Mersereau (2013)), we show that the optimal allocation policy gives priority to customers with higher profit margin. Interestingly, the optimal allocation quantity is independent of the customers’ future order volume and disruption risk does not matter. This is counter-intuitive, because it is natural to expect that big
customers with high potential future volume should get priority. Furthermore, loyalty may be prioritized or dis-prioritized, depending on the customers’ sensitivity to service level. We then show that the optimal acquisition policy is an add-up-to policy for each customer type. The optimal add-up-to levels indicate the desirable customer base portfolio. We show that these levels are increasing in demand uncertainty, implying the firm should focus more on the acquisition strategy to diversify customer demand risks when demand are more uncertain. Surprisingly, these levels are not necessarily increasing in the customers’ disruption risks. This is because the disruption risk not only increases the customer availability risks but also decreases the customers’ average demand and therefore its value for the company.

We next consider a finite horizon setting. This is suitable for products with short life cycles in rapidly changing market. Here, because of the end of horizon effect, the problem is no longer stationary. As a result, the independence of the optimal policy over the future demand volume disappears. In particular, customers with higher future demand volumes do receive priority. These priority decreases as the length of the planning horizon increases.

The remainder of this paper is organized as follows. In § 4.2, we briefly review the literature. § 4.3 introduces the components of our model, § 4.4 derives the supplier’s optimal capacity allocation and customer base policy under an infinite horizon and analyze the the impact of the disruption risk, retention risk, as well as the customer demand size and uncertainty on the supplier’s optimal strategy. § 4.5 analyze the two-period problem to investigate the effect of short life cycle. In § 4.6 , we discuss the insights and summarize the problem.

4.2 Literature Review

Our paper is primarily related to four bodies of literature: supply disruption risk management, service level dependent demand, customer acquisition and retention
balance, and optimal capacity allocation. Supply disruption risks have attracted significant attention in operations management in recent years. Many different tools and strategies for mitigating the risk of supply disruptions have been examined, including multi-sourcing (Babich et al. (2007); Dada et al. (2007); Federgruen and Yang (2008) & Federgruen and Yang (2009)), alternative supply and back up resources (Tomlin (2006); Yang et al. (2009)), and supply flexibility (Tomlin & Wang 2005). In this category, the most related work to ours is by Chaturvedi et al. (2014), who consider a buyer’s optimal order quantity allocation among multiple suppliers to retain the suppliers and maximize the total discounted expected profit. In contrast to these studies which focus on the risks from the supplier side and consider the buyer’s strategy to improve the supplier availability, we consider the mirror image problem from the supplier’s prospective to hedge against its customer risks.

In the operations literature, the phenomenon that the current poor service level can damage the subsequent period’s demand was first captured by Schwartz (1966), who considers a firm’s inventory problem when the current stock-out influences future’s demand. This line of analysis has been further extended to more general model setting by several scholars, including Schwartz (1970), Fergani’s (1976), and Robinson (1990). Liberopoulos et al. (2010) quantify the unit stock-out penalty cost in the economic context. Several recent works incorporate competition and customer forgiveness investment. Hall and Porteus (2000), Liu et al. (2007) and Gaur and Park (2007) consider joint inventory and pricing decisions when a customer experiencing a stock-out can switch/defect to a competitor. Olsen and Parker (2008) consider joint advertising and periodic-review inventory decisions, assuming that a customer experiencing a stock-out has a possibility to get into a non-purchasing "latent" state and the company can win back the customer by advertisement. The most related work of this stream to ours is Aflaki and Popescu (2014). These authors consider
a firm’s dynamic problem to customize service to maximize its long-term value derived from one customer, with a focus on the structure of optimal service policy and impact of customer loyalty, memory, etc. While they decide service level for a single customer with no capacity limit, we consider customer service level choice for a pool of customers under the capacity limit constraint.

In marketing, there is abundant literature on customer relationship management; see Gupta and Lehmann (2008) and Reinartz and Venkatesan (2008) for surveys of the literature on this topic. The works mostly related to our paper are those consider the tradeoff between acquisition and retention strategy. For example, Berger and Bechwati (2001) consider the optimal budget allocation among the acquisition and retention to maximize the customer equity. Reinartz et al. (2005) investigate the problem using empirical data to maximize the firm’s long-run profit. They show that the successful strategy requires sufficient investment in both area. They also find that firms usually underinvest in customer retention. Dong et al. (2011) consider agent incentive design for acquisition and retention strategies when over acquisition has negative impact over retention. Our paper also consider the spoiling effect of acquisition over retention, but focus on the interplay of operational factors such as customer risks and supplier’s capacity utilization.

Our analysis of capacity allocation policy is related to the optimal capacity allocation literature in supply chain management. Cachon and Lariviere (1999) consider a two-period capacity allocation problem for a supplier with finite capacity facing demand from its homogenous customers and evaluate allocation scheme “turn-and-earn”. They show that this allocation scheme increases the supplier’s profit by incentivizing the customers to order more. Lu and Lariviere (2012) extend the above results to a more general infinite horizon setting. Similar to ours, Adelman and Mersereau (2013) consider a capacity allocation problem when the customers adjust its order quantity based on its past fill rate. While Adelman and Mersereau (2013)
develop a solution technique assuming an exponential smoothing model of customer goodwill dynamics, we assume customer service memory is Markovian and focus on the interactions of acquisition and retention strategies with presence of customer disruption risks. Different from Adelman and Mersereau (2013), who propose a greedy myopic policy which gives absolute priority to customers with higher profit margin, we derive an exact optimal capacity allocation policy and characterize the customers' priority as a combination of factors such as profit margin, demand volume, loyalty.

To the best of our knowledge, our work is the first to study customer base management in the supply chain literature taking into account various types of customer risks and finite production capacity.

4.3 Model

4.3.1 Supply Chain Structure

We consider a supply chain consisting of a supplier serving a group of business customers over time, as shown in Figure 4.1. The supplier has a finite production capacity $K$ in each period to make to order. Because of technological and knowledge requirement for the production process, the supplier’s capacity is expensive to invest
and to expand. For instance, building new factories, investing in new equipments, and hiring skilled employees are time-consuming, which cannot be done in short time. Thus, for the supplier, left-over capacity is costly and its goal is to maximize expected profit by optimally managing its capacity utilization.

There are \( k \) different types of customers; customers of the same type are homogeneous. We use \( i \) to denote the type index and \( j \) the customer index. Let \( n_i^t \) denote the number of type-\( i \) customers in period \( t \). The order quantity from the \( j^{th} \) type-\( i \) customer in period \( t \) is a random variable, which follows the same distribution as random variable \( D_i^t \). Let \( D^t = (D_1^t, ..., D_k^t) \) represent the demand vector in period \( t \). To focus on analyzing customer base and derive long-term stationary analysis, we assume that the demand vector \( D^t \) is independent over time \( t \), although demand from type \( i_1 \) customers \( D_{i_1}^t \) and type \( i_2 \) customers \( D_{i_2}^t \) are not necessarily independent. Without loss of generality, we assume that the type \( i \) customers have larger order volume than type \( i - 1 \) customers, i.e. \( D_1 \leq_{st} D_2 \leq_{st} ... \leq_{st} D_k \). The supplier charges the same unit price to type \( i \) customers. In correspondence to the stochastically higher order quantity, big customers usually are able to drive lower unit price due to economy of scale effect and large bargaining power. Thus we assume that the supplier’s profit margin \( \pi_i \) from type-\( i \) customers satisfy \( \pi_1 \geq \pi_2 \geq ... \geq \pi_k \).

4.3.2 Customer Risk Profile

The supplier encounters three different types of demand risk from the customers. Firstly, the customers are subject to disruption risk, triggered by external destructive events, which would temporarily cut the customers’ orders. We assume that for any type \( i \) customer, there is a constant possibility \( p_i \) so that it will disrupt in period \( t \). Let \( a_{ij}^t \) denote the availability state of the \( j^{th} \) type \( i \) customer in period \( t \), with \( a_{ij}^t = 1 \) implying that the customer \( j \) in type \( i \) is in the normal state in period \( t \), \( a_{ij}^t = 0 \) means the disruption state. Then \( a_{ij}^t \sim \text{Bernoulli} (1 - p_i) \). In the normal
state, the customer will submit order quantity \( d_{ij} \); however, in the disruption state, the customer stops operation and therefore does not place any order.

Secondly, the customers have some probability to quit the supplier permanently due to various reason. For example, the customers may switch to other service provider or choose to vertically integrate this operation if they are not satisfied with their service. Or it may permanently cease the production due to the market strategy change. Or the customer may go bankrupt due to financial reasons. Here, we call this internal risk “customer retention risk”. This risk is permanent in the sense that once the customer leaves, the supplier lose its demand permanently. However, this risk can be hedged by providing good service level. This is because when the customers receive better service, they shall have lower incentive to switch or integrate and probability to go bankrupt due to better operational position. To model the retention risk and incorporate its dependency over service level, we let \( s_{ij}^t \) represent the customer’s retention state with \( s_{ij}^t = 1 \) implying that the customer \( j \) in type \( i \) will stay for period \( t + 1 \), \( s_{ij}^t = 0 \) means the customer will permanently leave the supplier at the end of period \( t \). Similarly, we assume \( s_{ij}^t \sim \text{Bernoulli}(\alpha_i(x)) \) with probability \( \alpha_i(x) \) that the type \( i \) customers stay and probability \( \bar{\alpha}_i(x) = 1 - \alpha_i(x) \) the customers leave. The retention probability \( \alpha_i(x) \) is increasing and concave function in service level \( x \). The concept of consumer retention risk is not new. There has been extensive literature in marketing analyzing the retention probability functions \( \alpha(x) \). See Fader and Hardie (2009) for a review of the retention probability models. Common parametric retention functions in the literature includes: (1) logistic retention function \( \alpha_i(x) = \frac{1}{1+e^{-ax}}, a \geq 0 \) (Rust et al. (2004)); (2) exponential retention function \( \alpha_i(x) = 1-e^{-ax}, a \geq 0 \) (Berger and Nasr (1998)); (3) power retention function \( \alpha_i(x) = a + bx^c \), when \( a, b \geq 0, a + b \leq 1, 0 \leq c \leq 1 \). Here the retention probability function \( \alpha_i(x) \) also captures the customers’ loyalty towards the supplier. For example, in the logistic retention function, parameter \( a \) represents the customer’s loyalty factor. When \( a \)
increases, the probability to retain $\alpha_i(x)$ also increases given the same service level $x$ and we say the customer is more loyal to the supplier. Thirdly, the supplier also suffers from routine demand uncertainty, inherited from the customers’ random order $D_i$.

4.3.3 Supplier’s Problem

Facing all these types of demand uncertainty, the supplier runs risk of too low total order from the customers and substantial unused left-over capacity. The supplier can hedge against this risk, by increasing its customer base to improve its capacity utilization and maximize its total profit. For example, the supplier can acquire more customers by spending on advertising, hiring more experienced sales managers, investing in production technique to improve quality, etc. However, the effort is costly and once a new customer is found, there is an extra transaction cost for doing business with this new customer. For example, the supplier may need to invest in learning the need of the customer, customize their production and machine set-up to the requirements of the customer, hire new employees to manage the transaction with the customer. For the simplicity of analysis, we assume that the cost for each additional type $i$ customer is constant. We call this additional cost the "acquisition cost" and denote it as $\ell_i$. Apparently, for different types of customers, the acquisition cost also varies. For example, the expense to attract and maintain a larger customer is usually higher than that of a small customer, not only because of the limited number of big customers, but also due to the competition from other suppliers. In reality, there may be limited number of type $i$ customers in the existing customer pool. Thus the cost to acquire an extra type $i$ customer is not constant, but increasing in the supplier’s existing type $i$ customers base. This more general case can be incorporated by assuming acquisition cost function $\ell_i(n)$ representing the cost of acquiring $n$ type $i$ customers from zero customer base. Obviously, $\ell_i(n)$ is increasing
and $\ell_i(n) - \ell_i(n-1)$ is increasing in $n$. The maximal customer pool $N_i$ for type $i$ customer can be captured by assuming $\ell_i(n) = \infty$, if $n \geq N_i$. Under this new acquisition cost function, most results still hold. To derive more insights, we focus on the linear acquisition cost.

While larger customer base is beneficial for improving the supplier’s short term capacity utilization rate, it also increases the chance of capacity shortage for the supplier and therefore hurts the service level the supplier is able to provide. In other words, although acquisition strategy is useful for diversification the demand risks from the customers, it reduces the customers’ satisfaction and increases customer defection rates due to poor fill rate. Furthermore, it brings extra operational dilemma for the supplier. When the total order from the customers exceeds the fixed capacity, the supplier needs to decide how to utilize its resources and this question is non-trivial. There is short-term profit taking for prioritizing the small customers with high unit profit margin. However, it might ruin the supplier’s relationship with the big customers and risk losing their future business.

In this paper, we investigate the problem under two different settings:

1. **Infinite horizon setting:** The setting is suitable for describing mature and well established markets where the products are commodity type with long life cycle. For example, several of the critical components for computers and laptops such as memory and solid state drives are produced in a mature market dominated by several powerful companies such as Samsung and Intel. These products as critical components for computer systems are widely used and have long life cycle.

2. **Finite horizon setting:** In rapidly changing market, the product life cycle could be very short. For example, most cell phones in US has average shelf life less than two years. As a result, most non-critical components for these cellphones
expire and go out of market quickly.

We examine the problem under both settings to better understand the optimal strategy and the impact of life cycle length over the strategy. In § 4.4, we investigate the problem under the infinite horizon setting while in § 4.5, we extend the analysis to finite horizon setting.

For simplicity of analysis, we do not incorporate the case when customers are strategic and adjust their order quantity based on the allocation scheme. This is also reasonable, since for most suppliers, their customer order and allocation data is rather confidential, which most customers usually have no access to. To focus on the capacity allocation problem, we also assume the supplier operates by make-to-order system. In other words, we do not consider the case when the supplier produce in advance for future demand.

4.4 Infinite Horizon Problem

4.4.1 Dynamic Programming Formulation

Under infinite horizon setting, we formulate the problem as a dynamic programming model with per period time discount parameter $\beta$. The events occur at the following order:

1. At the beginning of each period $t$, the supplier observes the number of each type customers $n_t^i$ and decides whether to acquire more customers to improve the capacity utilization. If the supplier decides to acquire, then how many $z_i^t = m_i^t - n_i^t$ customers of each type to acquire at unit acquisition cost $\ell_i$;

2. After that, the availability state $a_{ij}^t$ of the customers are realized and non-disrupted customers submit their orders $d_{ij}^t$ to the supplier simultaneously;

3. Observed the orders $d_{ij}^t$ from the customers and its capacity limit $K$, the supplier decide how much capacity $q_{ij}^t$ to allocate to each customer, to maximize
its total discounted expected profit;

4. According to the chosen allocation scheme, sale happens and the supplier’s profit is generated;

5. Observing the received service level \( x_{ij}^t = \frac{q_{ij}^t}{d_{ij}^t} \), customers decides to stay or quit with the retention probability \( \alpha_{ij}(x_{ij}^t) \) at the end of this period \( t \). The supplier’s customer pool is then updated with new number \( n_{i,t+1} \) of type \( i \) customers.

See Figure 4.2 for an illustration of timeline.

At each period, the supplier has two decisions to make: (1) stage 1: acquisition decision, the supplier has to decide how many customers to acquire \( z_i^t \) for each type; (2) stage 2: retention decision, when orders from the customers are realized, the supplier needs to decide the optimal allocation scheme. Let \( x_i^t = (x_{i,j}^t)_{j=1}^{n_i} \) be fill rate for type \( i \) customers and \( x^t = (x_i^t, \ldots, x_k^t) \) be fill rate in period \( t \). Then the supplier’s feasible allocation set in period \( t \) is

\[
\Omega^t = \left\{ x^t \mid 0 \leq x_{ij}^t \leq 1 : \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}^t a_{ij}^t \leq K \right\}
\]

satisfying the capacity limit constraint. When the customer is disrupted \( (a_{ij}^t = 0) \) and therefore submits no order, there is no dissatisfaction due to poor service level and under this case we assume \( x_{ij}^t = 1 \). Given the acquisition decision \( z_i^t \) and
retention decision $x_t$, the initial number of customers at the beginning of period $t$ is denoted as $n_t = (n_{t1}, n_{t2}, ..., n_{tk})$, which is updated as

$$n_{t+1}^i = \sum_{j=1}^{n_{t+1}^i + z_t^i} s_{ij}^t (a_{ij}^t, d_{ij}^t, x_{ij}^t)$$

in period $t + 1$. Here $s_{ij}^t$ is Bernoulli distributed with probability $\alpha_i (x_{ij}^t)$ to stay, which is endogenously determined by exogenous variables $a_{ij}^t$ and $d_{ij}^t$. Thus, the supplier’s problem can be formulated as following:

$$J_t (n_t) = \max_{m_t \geq n_t} E_{d,a} \left\{ \max_{i=1}^{m_t} \left( \pi_i \sum_{j=1}^{m_t} x_{ij}^t d_{ij}^t a_{ij}^t - \ell_i (m_t - n_t^i) \right) + \beta E_s [J_{t+1} (n_{t+1}^{t+1})] \right\}$$

The goal is to maximize the supplier’s total discounted expected profit and $J_t (n_t)$ is the optimal profit-to-go function when there is $n_t^i$ number of type $i$ customers at the beginning of period $t$.

### 4.4.2 Optimal Strategy Analysis

The dynamic programming of joint acquisition and retention decision is difficult to solve in its original. Further, the retention decision is endogenously affected by the acquisition decision. We address this challenge in three steps. First, we present the structure of the supplier’s optimal acquisition strategy. Second, based on the optimal acquisition strategy structure, we are able simplify the above dynamic programming problem and solve the optimal retention (capacity allocation) strategy. Third, the optimal total discounted expected profit function is simplified enough to solve for the optimal acquisition strategy.

#### Structure of Optimal Acquisition Strategy

We start by presenting the structure of optimal acquisition strategy,
Proposition 24. Given the initial number of customer $n_0^i = 0$, we can show that there exists $n_i^*$ so that the optimal acquisition policy is an add-up-to policy with the up-to level $n^* = (n_1^*, ..., n_k^*)$.

From Proposition 24, we know maintaining large customer base is not beneficial for the supplier. Instead, the optimal acquisition policy is a threshold policy. It is optimal for the supplier to maintain $n_i^*$ number of type $i$ customers in its customer base. Whenever $n_i$ drops below the target optimal $n_i^*$ due to customer quitting, the supplier should acquire $n_i^* - n_i$ type $i$ customers to make up for the loss. This result conforms with the existing results in marketing (Reinartz et al. (2005)) considering the balance between acquisition and retention strategy. It shows the optimal policy should balance the acquisition and retention strategy and too large or too small customer pool will hurt the supplier’s profit.

Optimal Retention Strategy

Given the structure of the optimal acquisition strategy is an add-up-to policy with stationary add-up-to levels and the customer order behavior is independent over time, we can decompose the infinite horizon problem into repeated stationary problem of one period. For the stationary process, we can suppress the superscript $t$ for the notations. Assuming the supplier starts the process with no customers, we can simplify the supplier’s total discounted expected function $J(n)$ when it operates the system with add-up-to level $n = (n_1, ..., n_k)$ as following:

$$J(n) = \frac{1}{1 - \beta}E_{d,a} \left[ \max_{\Omega} \sum_{i=1}^{k} \left\{ \sum_{j=1}^{n_i} \pi_i x_{ij} a_{ij} d_{ij} - \beta \ell_i \bar{a}_i (x_{ij}) \right\} \right] - \sum_{i=1}^{k} \ell_i n_i$$

Note that $J(n)$ consists of two terms: the first term is the total discounted expected sum of supplier’s profit for the infinite horizon, and the second term is the amortized cost of starting the customer base with $n_i$ type $i$ customers at the beginning of the
horizon. Given the customers’ availability \(a_{ij}\) and order \(d_{ij}\), the supplier needs to solve its operational level question and decides how to allocate its limited capacity \(q_{ij}\). This is equivalent to solve

\[
\max_{\Omega} \sum_{i=1}^{k} \left\{ \sum_{j=1}^{n_i} \pi_i x_{ij} a_{ij} d_{ij} - \beta \ell_i \bar{\alpha}_i (x_{ij}) \right\}
\]

When the total order \(\sum_{i=1}^{k} \sum_{j=1}^{n_i} a_{ij} d_{ij}\) is below the supplier’s capacity \(K\), there is no shortage and the supplier satisfies every customer’s order with \(q_{ij} = a_{ij} d_{ij}\). However, when capacity shortage occurs, the problem becomes more complicated. There is a tradeoff between short-term and long-term profit. Under this setting, we have

**Lemma 25.** When capacity shortage occurs \(\left( \sum_{i=1}^{k} \sum_{j=1}^{n_i} a_{ij} d_{ij} > K \right)\), there exists unique optimal allocation scheme \(x^*\) satisfying the Kuhn-Tucker first-order condition.

Lemma 25 guarantees the existence of unique optimal solution. Further, we want to investigate the supplier’s optimal allocation scheme.

**Proposition 26.** When capacity shortage occurs \(\left( \sum_{i=1}^{k} \sum_{j=1}^{n_i} a_{ij} d_{ij} > K \right)\), there exists \(\lambda\) so that the optimal fill rate

\[
x^*_{ij} = \begin{cases} 
0, & \text{if } \left(\alpha'_i\right)^{-1} \left( d_{ij} \frac{\lambda - \pi_i}{\beta \ell_i} \right) \leq 0 \\
\left(\alpha'_i\right)^{-1} \left( d_{ij} \frac{\lambda - \pi_i}{\beta \ell_i} \right), & \text{if } 0 < \left(\alpha'_i\right)^{-1} \left( d_{ij} \frac{\lambda - \pi_i}{\beta \ell_i} \right) < 1 \\
1, & \text{if } \left(\alpha'_i\right)^{-1} \left( d_{ij} \frac{\lambda - \pi_i}{\beta \ell_i} \right) \geq 1
\end{cases}
\]

Further, \(x^*_{ij}\) is decreasing in order quantity \(d_{ij}\), increasing in unit profit margin \(\pi_i\), unit acquisition cost \(\ell_i\) and retention sensitivity \(\alpha'_i()\) and independent of disruption risk \(p_i\) and demand volume \(D_i\).

Proposition 26 shows that when faced with capacity shortage, the priority of a customer is determined by measurement \(\left(\alpha'_i\right)^{-1} \left( d_{ij} \frac{\lambda - \pi_i}{\beta \ell_i} \right)\), which is a combination...
of the customers’ profit margin, current period demand volume, acquisition cost and retention risk. The higher the measurement is, the higher priority and fill rate the customer get when the supplier allocates its limited resource. It is intuitive that customers with higher unit profit margin $\pi_i$ should be prioritized, consistent with the results in Adelman and Mersereau (2013) and Aflaki and Popescu (2014). At the same time, the customers with higher cost $\ell_i$ to acquire should also be prioritized, since they are more expensive to acquire once lost. Interestingly, while we might expect customer with large order quantity $d_{ij}$ get prioritized, the result shows the opposite. This is because the supplier prefers large order only when it has capacity to fulfill and when capacity shortage occurs, large order only bring up difficulty for the supplier to satisfy the demand. When the total order exceeds its capacity limit, the supplier’s capacity is limited and its focus is to retain as many customers as possible using the limited capacity. In this case, since small order can be fulfilled with less resources, it is reasonable to give them priority.

Counterintuitively, the supplier does not prioritize the customers with high future demand volume $D_i$. Furthermore, while the literature (Adelman and Mersereau (2013)) suggests the supplier should emphasize the customers with lower demand uncertainty, our results suggest the demand uncertainty plays no role in the optimal capacity allocation decision. This is because the demand volume $D_i$ and uncertainty affects the profitability of that type of customers. Based on that, the supplier decides how many of this type customers to maintain in the customer base. Once this decision is made and shortage occurs, the supplier tries to balance the profit $\pi_i$ and customer leaving cost $\ell_i$, which is independent of future demand volume. Obviously, large and steadily ordering new customers are usually harder to find, leading to higher unit acquisition cost $\ell_i$. Therefore, large customers get prioritized because they are more expensive to acquire. The same logic applies to the finding that optimal capacity allocation is independent of customers’ disruption uncertainty $p_i$. 

96
Moreover, the analysis shows no clear trend of dependency between capacity allocation fill rate and customer retention function $\alpha_i(x)$. To illustrate this result further, let’s look at two examples of retention risk function: (1) $\alpha_i(x) = a + b_i x$, $a, b_i \geq 0$ and $a + b_i \leq 1$. In this example, customers with higher $b_i$ have higher probability to stay. Indeed, these customers are rewarded with higher fill rate when capacity shortage occurs, since $x_{ij}^*$ is increasing in $\alpha_i(x) = b$. (2) $\alpha_i(x) = a - b_i (1 - x)$, $a \geq b_i \geq 0$ and $a \leq 1$. In this example, customers with lower $b_i$ have higher probability to stay. However, these customers are not rewarded for their loyalty with higher priority, since $x_{ij}^*$ is increasing in $\alpha_i'(x) = b_i$. To further investigate the impact of customer loyalty over service level, we assume logistic and exponential retention function where parameter $a$ captures the customer’s loyalty. We have

**Proposition 27.** Under logistic or exponential retention functions, when the customer’s loyalty factor $a$ increases, its optimal fill rate $x_{ij}^*$ is first increasing then decreasing.

Together with Proposition 26, we show that more loyal customers does not necessarily have priority at the supplier. The customer’s optimal fill rate is not directly dependent on the magnitude of its retention probability $\alpha_i(x)$ or loyalty. Instead, it is dependent and increasing in the customers’ sensitivity to bad performance $\alpha_i'(x)$. When the customers are sensitive to bad performance, their retention probability drops quickly when the performance is bad and $\alpha_i'(x)$ is large. For the supplier with the fixed capacity, its goal is to maximize its customers’ retention. This can be mostly efficiently done by allocating the resources to the customers who are most sensitive to bad performance and whose retention probability increases most when their service level improves.
Optimal Acquisition Strategy

After we solve the operational question of capacity allocation, the optimal profit function is then further simplified. Then the more challenging problem is the strategic question of finding the exact optimal customer base size. To solve that, define

\[
G (\mathbf{m}) = E_d \left[ \max_{\Omega} \sum_{i=1}^{k} \left\{ \sum_{j=1}^{m_i} \pi_i x_{ij} d_{ij} - \beta \ell_i \bar{\alpha}_i (x_{ij}) \right\} \right]
\]

as the per-period profit function when \( \mathbf{m} = (m_1, \ldots, m_k) \) customers are not disrupted. Then through simplification, the supplier’s profit function \( J(n) \) given number of customer for each type \( \mathbf{n} \) can be rewritten as

\[
J (\mathbf{n}) = \frac{1}{1 - \beta} E_a \left[ G \left( \sum_{j=1}^{n_1} a_{1j}, \ldots, \sum_{j=1}^{n_k} a_{kj} \right) \right] - \frac{\beta}{1 - \beta} \sum_{i=1}^{k} \ell_i \bar{\alpha}_i (1) n_i p_i - \sum_{i=1}^{k} \ell_i n_i
\]

Therefore, optimizing the customer base size \( \mathbf{n}^* \) is equivalent to solve \( \max_{\mathbf{n}} J(\mathbf{n}) \).

Firstly, we obtain the following property of \( J(\mathbf{n}) \):

**Lemma 28.** When \( \alpha'(x) x \) is increasing, \( J(\mathbf{n}) \) is sub-modular and piecewise concave in \( \mathbf{n} \), i.e. for any non-negative integer vector \( \mathbf{n} \), the inequality \( J(n_i + 1, n_j, \mathbf{n}_{-i,j}) + J(n_i, n_j + 1, \mathbf{n}_{-i,j}) \geq J(n_i, n_j, \mathbf{n}_{-i,j}) + J(n_i + 1, n_j + 1, \mathbf{n}_{-i,j}) \).

From Lemma 28, we can see that the benefit of an additional type \( i \) customer is decreasing in the number of other type customers. This is quite intuitive since for supplier with finite capacity, the additional demand is less useful when it has already been in capacity shortage situation. Moreover, Lemma 28 guarantees that the optimal size for customers of each type \( \mathbf{n}^* \) is unique.

4.4.3 Sensitivity Analysis

The customer base size serves as an indicator for the supplier’s strategy preference: large \( n^* \) indicates that the supplier prefers acquisition strategy more to diversify risks.
and hedge against demand uncertainty; small $n^*$ suggests that the supplier emphasizes more on retention strategy to provide better service to its smaller customer base and reduce the retention risk. In order to compare the customer base size and derive more insights about the supplier’s strategy preference, we focus on the case when there is only one type of customer ($k = 1$). Specifically, we conduct the sensitivity analysis and analyze the impact of different market characteristics ($\pi, K, \ell, \beta$) on supplier’s strategy preference $n^*$.

It is obvious from Lemma 28 that when $k = 1$, the marginal value of an extra customer $J(n+1) - J(n)$ is non-increasing in $n$ and the optimal add-up-to level $n^*$ can be characterized:

$$n^* = \min \{n \geq 0 : J(n+1) - J(n) \leq 0\}$$

Based on that, we can derive the following sensitivity results about $n^*$:

**Proposition 29.** When there is only one type of customers, the optimal size $n^*$ is increasing in the unit profit margin $\pi$, capacity limit $K$, time discount $\beta$, decreasing in unit customer acquisition cost $\ell$.

The results from Proposition 29 are intuitive. When the supplier makes larger unit profit from selling product, it becomes more expensive to have left-over capacity and the supplier should favor the acquisition strategy more to improve its capacity utilization rate and thus profit. In addition, when the supplier has larger capacity to satisfy demand, naturally it should acquire a larger customer pool to utilize the capacity. At the same time, when the cost to acquire new customers is larger, the relative value of having a new customer is smaller. Thus acquisition strategy becomes less preferable and the supplier should focus more on retention strategy and improve its service level and retain its existing customers.

More interestingly, we want to investigate the supplier’s strategy preference when its customers’ risk profile changes (disruption risk $p$, retention risk $\alpha(x)$ and demand
uncertainty $D$). Specifically, these different type of risks have distinguishing features and we are interested in examining whether these features lead to different strategy preference. First, we have

**Proposition 30.** *When the uncertainty of demand $D$ from the customer increases, i.e. $D$ increases in convex order, $n^*$ is increasing.*

The result is reasonable, suggesting that when there are more uncertainty from customers’ demand, the supplier has higher chance to probability for left-over capacity. As a result, the supplier should focus more on acquisition strategy to diversify the risks. See Figure 4.3 left for an example of the optimal customer base change\(^1\). In the figure, $D \sim \text{Uniform}[k, 2-k]$. When $k$ increases, $D$ decreases in convex order and the demand $D$ becomes less uncertain. As can be seen from the figure, the optimal customer base size $n^*$ also decreases since there is lower need to maintain large customer base for left-over capacity.

**Proposition 31.** *When the disruption risk $p$ increases, there exists $p^*$ so that $n^*$ is increasing when $p \leq p^*$, decreasing when $p > p^*$.*

Proposition 31 provide a counterintuitive result, which suggests that it is not necessarily optimal for the supplier to always increase its customer base when the disruption risk from the customers becomes larger. See Figure 4.3 right for an example of the optimal customer base size\(^2\). As we can see, the optimal customer base increases in the disruption risk when the disruption risk is below 80% and decreases in the disruption risk when it is above 80%. This is because the disruption risk increases not only the demand uncertainty $pD$, it also reduces the average order size the customer places and therefore reduces the value of a customer. Based on this, when the disruption risk increases but still below a certain threshold $p^*$, the effect

\(^1\) $K = 4, \ p = 0.2, \ \pi = 2, \ \alpha(x) = x, \ \ell = 0.5, \ \beta = 0.95$

\(^2\) $K = 4, \ D \sim \text{Uniform}[0,1], \ \pi = 2, \ \alpha(x) = x, \ \ell = 3, \ \beta = 0.95$
of demand uncertainty incline is dominant. Thus, the supplier is optimal to acquire more customers to diversify the risks, just as when demand $D$ becomes more uncertain. However, when the disruption risk is above the threshold $p^*$, the effect of decreasing customer value becomes dominant. Consequently, the supplier has lower incentive to invest in acquiring new customers and chooses to focus more on retention strategy and improve the service level by reducing its customer base. In summary, the preference over acquisition strategy is only optimal when the disruption risk is relatively small.

**Proposition 32.** Assume the retention probability function is of form $\alpha(x, a)$ where $\alpha(x, a)$ is increasing in the customer’s loyalty factor $a$, the optimal customer size $n^*$ is also increasing in $a$ if $\frac{\partial \alpha(x, a)}{\partial a} > 0$.

The analysis shows that under the constraints that higher retention probability also means higher sensitivity to bad performance, when there is lower risk from the customers’ retention, the supplier should maintain a larger customer base to diversify the disruption and demand uncertainty risk. Interestingly, this suggests that only lower retention risk from the customers does not guarantee that the optimality of larger customer base, which also coincides with the result from the optimal capacity allocation.
4.5 Finite Horizon Problem

In the main model, we consider the problem under the infinite horizon setting. This setting is suitable for companies selling product with long life cycle in mature and well-established market. However, in some rapidly changing market, the product life cycle could be very short. To investigate the problem under the new setting and examine the impact of product life cycle, we consider a modified version of our problem under two periods.

4.5.1 Dynamic Programming Formulation

Under two-period setting, the problem can be similarly formulated as

\[
J_1(n^1) = \max_{m_i^1 \geq n_i^1} E_{d,a} \left\{ \max_{\Omega^1} \sum_{i=1}^{k} \left( \pi_i \left( \sum_{j=1}^{m_i^1} x_{ij}^1 d_{ij}^1 a_{ij}^1 - \ell_i (m_i^1 - n_i^1) \right) \right) + \beta E_s \left[ J_2(n_1^2, \ldots, n_k^2) \right] \right\}
\]

\[
J_2(n^2) = \max_{m_i^2 \geq n_i^2} E_{d,a} \left\{ \max_{\Omega^2} \sum_{i=1}^{k} \left( \pi_i \left( \sum_{j=1}^{m_i^2} x_{ij}^2 d_{ij}^2 a_{ij}^2 - \ell_i (m_i^2 - n_i^2) \right) \right) \right\}
\]

Here \( J_i(n^i) \) is the optimal profit-to-go function from period \( i \). In period 2, the planning horizon is ending and thus the optimal profit \( J_2(n^2) \) contains only the current period’s profit. One thing to note is that, for some start-up companies in fast changing market, it is not clear that whether the company can make to the next round. This case can be captured by assigning a certain probability \( \rho \) that the company can survive to the next period and letting \( \beta = \gamma \rho \) where \( \gamma \) is the new per period time discount.

To examine the difference of the optimal policy and derive more insights, we consider the problem when there are only one or two types of customers, i.e., \( k = 1 \) or \( k = 2 \).
4.5.2 Optimal Strategy for One Type of Customers \( (k = 1) \)

The analysis of one type of customers allows us to compare the optimal customer base size under infinite horizon setting and two periods setting. This further gives us a proper way to derive the strategy preference between acquisition and retention under different settings. We solve the dynamic programming problem backwards and have:

**Proposition 33.** Given the initial number of customer \( n^1 = 0 \), there exists \( n^{i*} \) in each period \( i \) so that the optimal acquisition policy is an add-up-to policy with the up-to level \( n^{i*} \). Moreover, \( n^{1*}(\gamma \rho) \) is increasing in the probability \( \rho \) with \( n^{1*}(0) = n^{2*} \).

As can be seen from Proposition 33, the optimal customer base policy is still an add-up-to policy whenever the number of customers drop below the up-to level, the supplier should acquire more customers to it. Due to the end of game effect, these up-to levels are no longer stationary. They are in fact time-dependent, and decreases as the period approaches the end. The reason is that when the period approaches the end, the value of a new customer decreases. In addition, for the start-up setting, when there is higher probability that the company will make to next round, the company should invest more on attracting new customers. This is because, the value of a customer increases when the company is more likely to survive by the same logic.

Next, we consider the optimal capacity allocation policy:

**Proposition 34.** In period 2, any capacity allocation scheme satisfying the maximal amount of demand is optimal. In period 1, the optimal fill rate \( x_j^{1*} \) is decreasing in the order volume \( d_j^1 \). Moreover, the optimal capacity allocation scheme is maximize \( n^2 = \sum_{j=1}^{m^1} s_j (\alpha (x_j^1)) \) in stochastic order.

At the last period, the supplier does not care about the customers’ retention anymore due to the end of game effect. Therefore, any allocation policy satisfies as
much demand as possible is optimal. Consistently, we have in period 1, the customers with higher current order volume is de-prioritized in terms of fill rate. Interestingly, unlike under infinite horizon problem where the optimal retention strategy is to maximize the expectation of number of customers retained

$$\max \mathbb{E} \left[ \sum_{j=1}^{m^1} s^1_j \left( \alpha \left( x^1_j \right) \right) \right] \iff \max \sum_{j=1}^{m^1} \alpha \left( x^1_j \right)$$

Under the finite horizon setting, the optimal allocation policy is actually to maximize the number of customers retained itself, because the optimal acquisition policy is no longer stationary. Although this is not equivalent, based on the results from (Boland and Singh (2006)), these two have similar implications of maximizing customer retention. At last, as long as $\beta > 0$, the optimal capacity allocation policy is the same. In other words, as long as the start-up company has positive probability to make to the second round, it is always optimal to retain as many customers as possible.

4.5.3 Optimal Strategy for Two Types of Customers ($k = 2$)

The analysis of optimal strategy for two types of customers allows us to derive insights about the optimal retention (capacity allocation) policy. When there are two types of customers, i.e., $k = 2$, similarly we solve the optimal customer policy backwards.

**Proposition 35.** In period 2, the optimal capacity allocation strategy is to absolutely prioritize customers with larger profit margin. In period 1, the optimal fill rate $x^1_{ij}$ is decreasing in the order volume $d^1_{ij}$ and $p_i$, increasing in $\ell_i$, $\pi_i$, and future demand volume $D^2_i$.

Intuitively, at the last period, the supplier has no incentive to care about future profit. Therefore, the optimal capacity allocation scheme in period 2 is a greedy
policy, which is to maximize the supplier’s current period’s profit and prioritize the customers with larger profit margin absolutely. Namely, given the customers of type 1 has larger profit margin, the supplier always satisfy their orders first, then satisfy the orders of type 2 customers.

Consistent with the infinite horizon setting, the customers with higher profit margin, acquisition cost and sensitivity to service level and lower current period order volume is prioritized with higher service level. Interestingly, the optimal retention strategy is no longer independent of the customers’ future demand volume or disruption risk. In other words, under the finite horizon setting, the supplier does give priority to customers with large future potential volume to retain them, in anticipation of higher future profit. This is also true for the customers with lower probability to be disrupted. The reason is because under the finite horizon setting, the optimal acquisition strategy is no longer stationary, due to the end of game effect. As a result, the problem can not be decomposed into repeating single period problem and the optimal acquisition and retention strategy can not be separated. Therefore, the optimal retention strategy is no longer myopic policy, where only the customers’ retention cost is minimized.

4.6 Insights and Conclusion

In this work, we study a firm’s strategy to deal with its business customers’ demand risks in a business-to-business setting. Its customers are differentiated in volume, profit margin, loyalty and risk profile. We consider three types of demand risks from the customers: first, the customers are subject to exogenous and temporary disruption risks triggered by disaster events such as natural disasters, factory fire, labor strike, etc. Second, the customers are also prone to retention risks if they are not satisfied with the firm’s service level. This risk is endogenous determined by service level and permanent since once the customers leave, the firm loses their
business forever. Third, even without the above two types of risks, the firm faces
the usual, but substantial, uncertainty in the customers’ order variability. To hedge
against the demand risks, the firm can acquire more customers to diversify and
increase its capacity utilization. However, this may lead to a lower service level
and therefore higher customers’ retention risk. To guild firms to effectively make
these decisions, we explore the following questions: How should the supplier balance
its strategy preference between acquisition for customers’ order diversification and
retention for high customer retention rate? How to optimally rationalize capacity
under capacity shortage?

We build up a dynamic programming model and investigate the optimal strategy
under both the infinite horizon and finite horizon setting. Under the infinite
horizon setting, we derive the explicit optimal allocation policy when the total order
from customers exceeds the capacity. Consistent with the literature (Adelman and
Mersereau (2013)), we show that the optimal allocation policy gives priority to cus-
tomers with higher profit margin. Interestingly, the optimal allocation quantity is
independent of the customers’ future order volume, which contradicts with our intu-
iton that big customers with high potential future volume get priority. Furthermore,
loyalty could be prioritized or dis-prioritized, depending on the customers’ sensitiv-
ity to bad service level. In summary, we show that the optimal capacity allocation
policy is a stationary policy, which calculate the customers’ priority based on a com-
bination of factors, such as profit margin, acquisition cost, current demand volume
and sensitivity to bad performance. However, the policy is also modified myopic, in
the sense the customers’ future demand volume and disruption risk does not matter.

Under the infinite horizon, for the optimal acquisition strategy, there exists op-
timal add-up-to level for each type of customers so that it is optimal for the firm to
always acquire customers up to these levels. These optimal add-up-to levels serve
as indicators for firms’ preference towards acquisition and retention strategy. These
add-up-to levels are increasing in the routine demand uncertainty, implying the firm should focus more on acquisition strategy to diversify the customers' demand risks when the customers' demand are more uncertain. Surprisingly, these add-up-to levels are not necessarily increasing in the customers' disruption risks. This is because the disruption risk not only increases the demand risks but also decreases the customers' average demand and therefore its value for the company. When the disruption risk is too high, the customers' value decreases too much. As a result, the supplier may favor less on acquisition strategy to enlarge the customer pool.

At last, we investigate the problem under finite horizon setting, to characterize the impact of short life cycle in a rapidly changing market. When the problem has only two periods, there are still add-up-to levels at each period that the supplier wants to maintain. However, these add-up-to levels are no longer stationary given the supplier has less incentive to acquire more customers as the product approaches the end of life cycle, given the life-time value of new acquired customer is low. As a result, the add-up-to levels are decreasing over time. Moreover, because of the end of game effect, the problem is no longer stationary. As a result, the independency of the optimal capacity allocation policy over the future demand volume disappears. In other words, customers with higher future demand volumes do get prioritized. So do the customers with lower probability to disrupt.
Appendix A

Appendix

Proof of Proposition 1.

Proof. If \( k_0 \geq k_1 \), for any effort set \((e_0, e_1)\), a new effort set \((0, e_0 + e_1)\) would result in a non-increasing cost \( \Pi (w, \varphi, e_0, e_1) \). Therefore \( e_0^* = 0 \). Since \( \alpha (x) \) is a decreasing and convex function, \( \Pi (w, \varphi, e_0, e_1) \) is concave with the optimal

\[
e_1^* = \begin{cases} 0 & \text{if } -\alpha' (0) \leq \frac{k_1}{L}, \\ (\alpha')^{-1} (-\frac{k_1}{L}) & \text{otherwise}. \end{cases}
\]

The analysis for \( k_0 < k_1 \) is similar, which leads to the result in the proposition. □

Proof of Proposition 2.

Proof. We first analyze the Tier 1 supplier’s optimal effort strategy given the manufacturer’s wholesale price \( w \) and effort level \( e_0 \). The profit function for the Tier 1 supplier is

\[
\Pi_1 (w, 0, e_0, e_1) = (w - c) d - k_1 e_1 - \alpha (e_0 + e_1) L_1 (w).
\]
Since $\Pi_1(w,0,e_0,e_1)$ is concave with respect to $e_1$, some $e_1^\ast (e_0,w)$ is optimal only if it satisfies the first order condition, $\alpha' (e_0 + e_1^\ast (e_0,w)) = - \frac{k_1}{L_1(w)}$. This implies $e_1^\ast (e_0,w) = \max \{ e_1^d(w) - e_0, 0 \}$. Moving next to the manufacturer’s effort level, note that either the manufacturer will exert some effort $e_0 < e_1^d(w)$ or some effort $e_0 > e_1^d(w)$. In the former case, the manufacturer will achieve the same aggregate effort level $e_1^d(w)$ for any $e_0$, hence the manufacturer will choose $e_0 = 0$ to minimize cost. In the latter case, the manufacturer will induce the Tier 1 supplier to exert zero effort, leading to an optimal (manufacturer and aggregate) effort level equal to $e_0^d(w)$. Define $T = (\alpha (e_1^d(w)) - \alpha (e_0^d(w))) L_0(w) - k_0 e_0^d(w)$ to be the difference in profit under these two extreme strategies. $e_0^d = 0$ if $T < 0$ and $e_0^d = e_0^d(w)$ if $T \geq 0$. Let $\bar{w}$ satisfy $\frac{k_0}{L_0(w)} = \frac{k_1}{L_1(w)}$. When $w > \bar{w}$, $\frac{k_0}{L_0(w)} > \frac{k_1}{L_1(w)}$ and $T < 0$. When $w \leq \bar{w}$, $\frac{k_0}{L_0(w)} < \frac{k_1}{L_1(w)}$ and $\alpha (e_0^d(w)) < \alpha (e_1^d(w))$. Taking the derivative of $T$ over $w$, we have $\frac{\partial T}{\partial w} < 0$ when $w \leq \bar{w}$. Therefore, there exists a threshold wholesale price $w_s$ so that when $w \leq w_s$, $T \geq 0$, yielding the result in the proposition. Lastly, to test the sensitivity of $w_s$ to the model parameters, consider the monotonicity of $T$. Since $T$ is decreasing in $a_1$ and $w$ when $w \leq \bar{w}$, $w_s$ is decreasing in $a_1$. Similarly, it follows that $w_s$ is increasing in $a_0$ and $k_1$, decreasing in $k_0$.

Proof of Proposition 3.

Proof. The manufacturer’s optimal profit under the control strategy is

$$\max_{e \leq w \leq w_s} (p - w) d - k_0 e_0^c (w) - \alpha (e_0^c (w)) L_0(w)$$

$$s.t. \ (w - c) d - \alpha (e_0^c (w)) L_1(w) \geq r$$

Manufacturer profit is decreasing in $w$. Thus the manufacturer’s optimal control strategy price is the minimal price satisfying the Tier 1 supplier’s participation constraint, which is defined to be $w^c$. The manufacturer’s optimal profit under the
delegation strategy is

\[
\max_{w_s \leq w \leq p} (p - w) d - \alpha \left( e_1^d (w) \right) L_0 (w)
\]

s.t. \( (w - c) d - k_1 e_1^d (w) - \alpha \left( e_1^d (w) \right) L_1 (w) \geq r \)

The Tier 1 supplier’s profit is increasing in the wholesale price, thus there exists a threshold \( \hat{w} \) such that the participation constraint is satisfied for all \([\hat{w}, p]\). Next, we consider the manufacturer’s strategy preference. For \( k_1 \), the optimal profit under control strategy is independent of \( k_1 \). However, Tier 1 supplier profit is decreasing in \( k_1 \), therefore the feasible set \([\hat{w}, p]\) is shrinking (i.e., \( \hat{w} \) is growing). At the same time, \( w_s \) increases in \( k_1 \) (per the proof of Proposition 2) and the manufacturer profit is decreasing in \( k_1 \). Thus the optimal profit under a delegation strategy is decreasing in \( k_1 \), yielding part (iii) of the proposition. The result for \( k_0 \) follows similarly, yielding part (i).

To see the impact of \( r \) on the manufacturer’s strategy, consider the price that maximizes the manufacturer’s profit under the delegation strategy, which we denote as \( w_1 \). When there are multiple maximizers, set \( w_1 \) be be the largest one. When \( r \leq \Pi_1^d (w_1) \), the manufacturer’s optimal delegation strategy price is \( w^d (r) = w_1 \) and the manufacturer’s delegation strategy profit is independent of \( r \). At the same time, the manufacturer’s profit under control strategy decreases with \( r \) (which leads to a larger wholesale price \( w^c \)), leading to part (v) when \( r \leq \Pi_1^d (w_1) \). When \( r > \Pi_1^d (w_1) \), if \( \Pi_0^d (w) \) is decreasing on \([w_1, p]\), the manufacturer’s optimal delegation strategy price \( w^d (r) = \hat{w} (r) \) is increasing in \( r \) and the participation constraint is binding. The total supply chain profit under the delegation strategy is increasing due to higher system effort \( e_1^d (w^d (r)) \) while the total supply chain profit under control strategy is decreasing due to lower system effort \( e_0^c (w^c (r)) \). At the same time, the Tier 1 supplier receives the same reservation profit \( r \) under the control and delegation strategies. Thus, the difference of the manufacturer’s profit under delegation strategy and con-
trol strategy is also increasing. If $\Pi_0^d(w)$ is not monotone on $[w_1, p]$, i.e., there are other maxima of $\Pi_0^d(w)$ on the range $(w_1, p)$, find the maximum with the highest manufacturer profit and denote it as $w_2$. Define $b$ as the maximum wholesale price such that $\Pi_0^d(w)$ is decreasing on $(w_1, b)$. It must be true that $\Pi_0^d(w_2) \geq \Pi_0^d(b)$, otherwise there exists a maximum to the right of $b$ in which the manufacturer’s profit is larger than $\Pi_0^d(w_2)$, which conflicts with the definition that $w_2$ is the highest maximum point. Define $a$ as the maximum wholesale price such that $\Pi_0^d(w)$ is decreasing on $(w_1, w_2)$ with manufacturer’s profit larger than $\Pi_0^d(w_2)$. When $\Pi_1^d(a) > r > \Pi_1^d(w_1)$, the manufacturer’s optimal delegation strategy price $w^d(r) = \hat{w}(\gamma)$ is increasing in $r$ and the participation constraint is binding. The total supply chain profit under delegation strategy is increasing due to higher system effort $e_1^d(w^d(r))$ while the total supply chain profit under control strategy is decreasing due to lower system effort $e_0^c(w^c(r))$. At the same time, the Tier 1 supplier receives the same reservation profit $r$ under control and delegation strategy. Thus, the difference of the manufacturer’s profit under delegation strategy and control strategy is also increasing. When $\Pi_1^d(w_2) \geq r \geq \Pi_1^d(a)$, the manufacturer’s optimal delegation strategy price $w^d(r) = w_2$ and the manufacturer’s delegation strategy profit stays the same. Similarly, the difference of the manufacturer’s profit under the delegation strategy and control strategy is increasing. When $r > \Pi_1^d(w_2)$, continue the previous process on $[w_2, p]$, and it follows that the difference of the manufacturer’s profit under the delegation strategy and control strategy is increasing. Thus if $r \leq r'$ and the delegation strategy is optimal under $r$, it is also optimal under $r'$, proving part (v). The results for $a_0$ and $a_1$ follow similarly, proving parts (ii) and (iv).

Proof of Proposition 5.
Proof. Under substitutable effort, the system exerts all effort from the low cost stage \((\alpha')^{-1}\left(-\frac{\min\{k_0,k_1\}}{L}\right)\). It is straightforward that the effort is increasing in \(\gamma, \beta\) and \(a_i\).

Proof of Proposition 6.

Proof. (i) Fixing the manufacturer’s strategy to delegation, the manufacturer’s optimization problem is

\[
\max_{c \leq w \leq p} (p - w) d - \alpha \left(e_1^d (w)\right) L_0 (w)
\]

\[
s.t. \ (w - c) d - k_1 e_1^d (w) - \alpha \left(e_1^d (w)\right) L_1 (w) \geq r
\]

The optimal system effort is \(e_1^d (w^d) = (\alpha')^{-1}\left(-\frac{k_1}{L_1(w^*)}\right)\). Following the same logic as before, we can show that \(e_1^d (w^d)\) is increasing in \(a_0\) and \(\beta\). When \(\gamma\) increases, \(\Pi_1^d (w)\) decreases, thus \(\hat{w}\) increases. Under \(\alpha (x) = \beta e^{-x}\), \(\Pi_0^d (w)\) is concave in \(w\) with\[
\frac{\partial \Pi_0^d (w)}{\partial w} = -d + \gamma d \frac{k_1 L}{L_1 (w)^2}
\]

Let \(w_1 (\gamma)\) denote the unconstrained optimizer of \(\Pi_0^d (w)\), which satisfies \(L_1 (w_1 (r))^2 = L k_1 \gamma\). The optimal delegation price is \(w^d (\gamma) = \max \{\hat{w} (\gamma), w_1 (\gamma)\}\). From the equation, we know \(L_1 (w_1 (r))\) is increasing in \(\gamma\), and \(e_1^d (w_1 (r))\) increasing in \(\gamma\). At the same time, \(\hat{w} (\gamma)\) is increasing in \(\gamma\) and \(e_1^d (\hat{w} (\gamma))\) is increasing in \(\gamma\). Given \(\hat{w} (\gamma)\) and \(w_1 (r)\) is continuous, \(e_1^d (w^d (r))\) is also increasing in \(\gamma\). Using the same method, we can show that \(e_1^d (w^d (a_1))\) is increasing in \(a_1\). Note here, we ignore the case when \(\frac{\partial \Pi_0^d}{\partial w} (p) > 0\) \((w_1 (\gamma) > p)\). This is because under this case, the manufacturer’s profit function \(\Pi_0^d (w)\) is increasing over \([c, w_1 (\gamma)]\) and the manufacturer has incentive give the Tier 1 supplier wholesale price \(w\) higher than \(p\), i.e. \(w^d (\gamma) = w_1 (\gamma) \geq p\). Under this situation, \(L_1 (p) = \gamma (p - c) d + a_1\) increases and \(e_1^d (w^d (r))\) still increases. In
addition, this leaves the manufacturer negative optimal profit, which violates our assumption that the manufacturer should have non-zero profit.

(ii) Fixing the manufacturer’s strategy to be the control strategy, the manufacturer’s optimization problem is
\[
\max_{c \leq w \leq p} (p - w) d - k_0 e_0^c (w) - \alpha (e_0^c (w)) L_0 (w)
\]
\[
s.t. \ (w - c) d - \alpha (e_0^c (w)) L_1 (w) \geq r
\]
Since the manufacturer’s profit is decreasing in \(w\), the optimal control strategy price is the minimal price \(w^c\) satisfying \(\Pi_1^c (w^c) \geq r\). The resultant effort is \(e_0^c (w^c) = (\alpha')^{-1} \left( -\frac{k_0}{L_0 (w^c)} \right) \). When \(a_0\) increases, manufacturer’s loss function \(L_0 (w)\) increases. Thus \(\Pi_1^c (w)\) is increasing and \(w^c\) is decreasing in \(a_0\). Thus the system effort \(e_0^c (w^c)\) is increasing, because \(L_0 (w^c) = \gamma (p - w^c) d + a_0\) is increasing. Similarly, when \(a_1\) increases, the Tier 1 supplier’s loss function \(L_1 (w)\) is increasing. Thus \(\Pi_1^c (w)\) is decreasing and \(w^c\) is increasing in \(a_0\). The system effort \(e_0^c (w^c)\) is decreasing, because \(L_0 (w^c) = \gamma (p - w^c) d + a_0\) is decreasing. Under the exponential risk function, \(e_0^c (w) = (\lambda')^{-1} \left( -\frac{k_0}{\beta L_0 (w)} \right) \). The manufacturer’s profit function is
\[
\max_{c \leq w \leq p} (p - w) d - k_0 e_0^c (w) + k_0
\]
\[
s.t. \ (w - c) d + \frac{k_0}{L_0 (w)} L_1 (w) \geq r
\]
Since \(\Pi_1^c (w)\) is independent of \(\beta\), \(w^c\) is independent of \(\beta\). The system effort \(e_0^c (w^c)\) is increasing in \(\beta\), since \(\beta L_0 (w^c)\) is increasing in \(\beta\). Similar analysis applies for \(\gamma\). \(\Box\)

**Proof of Proposition 7.**

*Proof.* When \(a_0 \leq \alpha_0\), the delegation strategy is optimal and \(w^*\) is increasing in \(a_0\). Thus when \(a_0 \leq \alpha_0\), the system optimal effort \(e_0^d (w^*) = (\alpha')^{-1} \left( -\frac{k_1}{L_1 (w^*)} \right) \) is increasing. When \(a_0 > \alpha_0\), the control strategy is optimal. The optimal price is
$w^c$ is decreasing and the system effort $e_0^c(w^c) = \left(\alpha'\right)^{-1}\left(-\frac{k_0}{L_0(w^c)}\right)$ is increasing. In order to prove that the optimal effort is increasing globally, we need to show that the optimal system effort is increasing even when the optimal strategy changes from delegation to control or vice versa. According to the Envelope Theorem, we have

$$\frac{\partial \Pi_0^c(w^c)}{\partial a_0} = -\alpha \left(\left(\alpha'\right)^{-1}\left(-\frac{k_0}{\gamma (p - c) d + a_0}\right)\right)$$

$$\frac{\partial \Pi_0^d(w^d)}{\partial a_0} = -\alpha \left(\left(\alpha'\right)^{-1}\left(-\frac{k_1}{\gamma (\bar{w} - c) d + a_1}\right)\right)$$

Here $w^d$ is the optimal price under delegation strategy. At $a_0$, we have $\Pi_0^c(w^c) = \Pi_0^d(w^d)$ according to the intermediate value theorem. If the system optimal strategy changes from delegation to control, we must have

$$\frac{\partial \Pi_0^c(w^c)}{\partial a_0} \geq \frac{\partial \Pi_0^d(w^d)}{\partial a_0}$$

$$\left(\alpha'\right)^{-1}\left(-\frac{k_0}{\gamma (p - w^c) d + a_0}\right) \geq \left(\alpha'\right)^{-1}\left(-\frac{k_1}{\gamma (w^d - c) d + a_1}\right)$$

Therefore, the optimal effort increases at $a_0$ and increases generally in $a_0$. However, $e^*$ is not monotone in $\gamma$, $a_1$, or $\beta$, as illustrated by the counterexamples in the discussion. \qed

**Proof of Proposition 8.**

*Proof.* To coordinate the supply chain, we need to incentivize the centralized optimal effort $\left(\alpha'\right)^{-1}\left(-\frac{k_1}{L}\right)$ from the Tier 1 supplier when $e_0 = 0$. Since $e^d_1(w, \varphi) = \left(\alpha'\right)^{-1}\left(-\frac{k_1}{\gamma (w - c) d + a_1 + \varphi}\right)$, we must have $\varphi = \gamma (p - w) d + a_0$. However, the Tier 1 supplier accepts contract if and only if its expected profit is greater than $r$. Since

$$\Pi^* = (p - c) d - k_1 \left(\alpha'\right)^{-1}\left(-\frac{k_1}{L}\right) - \alpha \left(\left(\alpha'\right)^{-1}\left(-\frac{k_1}{L}\right)\right) L \geq r$$

and $(c - c) d - k_1 \left(\alpha'\right)^{-1}\left(-\frac{k_1}{L}\right) - \alpha \left(\left(\alpha'\right)^{-1}\left(-\frac{k_1}{L}\right)\right) L < r$
according to intermediate value theorem, there exists some \( w \) so that the Tier 1 supplier’s profit exactly equals \( r \). Therefore, the Tier 1 supplier accepts the contract and the coordinating contract is guaranteed to exist. Moreover, the optimal wholesale price \( w \) is

\[
w^* = \frac{\alpha \left( (\alpha')^{-1} \left( -\frac{k_1}{L} \right) \right) L + k_1 (\alpha')^{-1} \left( -\frac{k_1}{L} \right) + r}{d} + c
\]

\[
= (1 - \theta^*) (p - c) + c
\]

with the optimal penalty \( \varphi^* = \gamma (p - w^*) d + a_0 \). Furthermore, \( w^* \) is increasing in \( k_1 \) with first order derivative \( (\alpha')^{-1} \left( -\frac{k_1}{L} \right) \), and therefore \( \varphi^* \) is decreasing in \( k_1 \). Similarly, it is straightforward to show \( w^* \) is increasing in \( a_0, a_1 \) and \( t \), and \( \varphi^* \) is decreasing in \( a_1 \) and increasing in \( a_0 \). To see the impact of \( \gamma \) on \( \varphi^* \), the first order derivative is \( \left( \theta^* - \gamma (\alpha')^{-1} \left( -\frac{k_1}{L} \right) \right) (p - c) d \). Thus \( \varphi^* \) is increasing when \( \theta^* \geq \gamma \alpha^* \) and decreasing in \( \gamma \) if \( \theta^* < \gamma \alpha^* \).

Proof of Proposition 9.

\( \square \)

**Proof.** To coordinate the supply chain, we need to incentivize the centralized optimal effort \( (\alpha')^{-1} \left( -\frac{k_0}{L} \right) \) from the manufacturer, with \( e_1 = 0 \). From the previous analysis, we know the manufacturer exerts zero or \( e_0^d (w, \varphi) = (\alpha')^{-1} \left( -\frac{k_0}{\gamma (p - w) d + a_0 - \varphi} \right) \) effort. And it exerts effort \( e_0^d (w, \varphi) \) effort only if \( \varphi \geq w - a_1 \). Therefore we must have

\[
\frac{k_0}{\gamma (p - w) d + a_0 - \varphi} = \frac{k_0}{L}
\]

\[-\varphi = \gamma (w - c) d + a_1 \]

When \( -\varphi = \gamma (w - c) d + a_1 \), the Tier 1 supplier encounters zero loss \( L_1 (w) = 0 \) from non-compliance and therefore has no incentive to exert effort, i.e.\( e_1 = 0 \). In addition, there exists \( w^* \in [c, p] \) so that the Tier 1 supplier receives reservation profit \( r \),

\[
(w^* - c) d = r \iff w^* = \frac{r}{c} + d
\]

115
Given
\[(p - c) d \geq r + k_0 \left( \alpha' \right)^{-1} \left( -\frac{k_0}{L} \right) + \alpha \left( \left( \alpha' \right)^{-1} \left( -\frac{k_0}{L} \right) \right) L\]

and \((c - c) d < r\)

At last, we need to show that \(w^* > w_s\), i.e. given \(\varphi^*\) and \(w^*\), the manufacturer has higher profit under control strategy than that under delegation strategy. Under control strategy, manufacturer has profit \((p - c) d - r - k_0 \left( \alpha' \right)^{-1} \left( -\frac{k_0}{L} \right) - \alpha \left( \left( \alpha' \right)^{-1} \left( -\frac{k_0}{L} \right) \right) L\).

Under delegation strategy, the manufacturer has profit \((p - c) d - r - \alpha (0) L\). Apparently, the manufacturer has larger profit under control strategy. Therefore, the optimal coordinating contract has \(\varphi^* = -a_1 - \gamma r\) and \(w^* = \frac{r}{c} + d\).

\[\square\]

**Proof of Proposition 10.**

*Proof.* Under the centralized setting, the supply chain profit is
\[(p - c) d - k_0 e_0 - k_1 e_1 - \alpha (e_0 e_1) L\]

The optimal solution could follow two cases: (1) \(e_0^* = 0, e_1^* = 0\); (2) \(e_0^* > 0, e_1^* > 0\).

In the latter case,
\[-k_0 - \alpha' (e_0^* e_1^*) e_1^* L = 0\]
\[-k_1 - \alpha' (e_0^* e_1^*) e_0^* L = 0\]

Thus \(e_0^* = \frac{k_1}{k_0} e_1^*\). The problem is equivalent to solving
\[
\max_{e_0 \geq 0} \Pi (w, \varphi, e_0) = (p - c) d - 2k_0 e_0 - \alpha \left( \frac{k_0}{k_1} \frac{e_0^2}{e_0} \right) L
\]

When \(e_0 \to \infty\), \(\Pi (w, \varphi, e_0) \to -\infty\). There exists a threshold \(a\) such that when \(e_0 \geq a\), \(\Pi (w, \varphi, e_0) \leq \Pi (w, \varphi, 0)\). The maximizer of a continuous function over a closed set \([0, a]\) is guaranteed to exist. Hence, if \(e_0^* > 0\), it satisfies \(\alpha' \left( \frac{k_0}{k_1} \frac{e_0^*}{e_0} \right) = -\frac{k_1}{e_0 L}\).
Define $\Omega^+$ to be the set of parameter values for which
\[
\left\{ \max_{e_0 > 0} - 2k_0e_0 + \left( \alpha (0) - \alpha \left( \frac{k_0}{k_1} e_0^2 \right) \right) L > 0 \right\}.
\]
If the system is in set $\Omega^+$, the optimal effort $e_0^* > 0$; hence, if the set $\Omega^+$ expands with respect to a particular parameter, then parameter values that enter the set result in “increased” effort (i.e., move from zero optimal effort to non-zero effort). Specifically, we consider the sensitivity of set $\Omega^+$ to the supply chain loss function $L$. It is straightforward that when $L$ increases, $\Omega^+$ expands. Moreover, consider the sensitivity of $e_0^*$ with respect to $L$. Taking the derivative over $L$, we have
\[
\left( 2\alpha^" \left( \frac{k_0}{k_1} e_0^2 \right) \frac{k_0}{k_1} e_0^* + \alpha^' \left( \frac{k_0}{k_1} e_0^2 \right) \right) \frac{\partial e_0^*}{\partial L} = \frac{k_1}{L^2}
\]
Based on the optimality of $e_0^*$, we know
\[
\frac{\partial^2 \Pi (w, \varphi, e_0)}{\partial^2 e_0} \bigg|_{e_0=e_0^*} = -2\frac{k_0}{k_1} L \left( 2\alpha^" \left( \frac{k_0}{k_1} e_0^2 \right) \frac{k_0}{k_1} e_0^* + \alpha^' \left( \frac{k_0}{k_1} e_0^2 \right) \right) \bigg|_{e_0=e_0^*} \leq 0.
\]
Thus $\frac{\partial e_0^*}{\partial L} > 0$, i.e., $e_0^*$ is increasing in $L$. When the external pressure increases from the centralized supply chain ($\gamma, \beta, a_i$ increases), $L$ increases. Therefore, the equilibrium effort level $\frac{k_0}{k_1} e_0^*^2$ is increasing.

Under a decentralized supply chain with $\alpha(x) = \beta e^{-x}$, we need to solve the manufacturer’s optimal strategy. First, given the manufacturer’s effort level $e_0$ and wholesale price $w$, consider the Tier 1 supplier’s optimal effort strategy $e_1^* (e_0, w)$. The Tier 1 supplier’s profit is $(w - c) d - k_1 e_1 - \alpha (e_0 e_1) L_1 (w)$, hence we have $e_1^* (e_0, w) > 0$ if and only if $\alpha^' (e_1 e_0) e_0 = -\frac{k_1}{L_1 (w)}$. Since $\alpha^' (e_1 e_0) e_0$ is increasing in $e_1$, this holds if and only if $\alpha^' (0) e_0 < -\frac{k_1}{L_1 (w)}$, i.e., $e_0 > e_0^* (w) = -\frac{k_1}{\beta L_1 (w)}$. In this case, $e_1^* (e_0, w) = \left( \frac{\alpha^'}{\alpha^"} \right)^{-1} \left( \frac{k_1}{L_1 (w)} \right) \frac{1}{e_0}$. If $e_0 \leq e_0^* (w)$, the optimal $e_1^* (e_0, w) = 0$.

Next, we consider the manufacturer’s optimal effort $e_0^*$ given $w$. When $e_0 \geq e_0^* = \frac{k_1}{\beta L_1 (w)}$, the manufacturer’s profit function is $(p - w) d - k_0 e_0 - \frac{k_1}{L_1 (w) w} L_0 (w)$.
Optimizing over \( e_0 \geq e_0^* \), the optimum is \( \sqrt{\frac{k_1L_0(w)}{k_0L_1(w)}} \) if \( L_0(w)L_1(w) > \frac{k_1k_0}{\beta^2} \). Otherwise, it is \( e_0^* \) if \( L_0(w)L_1(w) \leq \frac{k_1k_0}{\beta^2} \). Further, we consider the global optimum \( e_0^* \)

\[
\max \left\{ (p - w) d - \alpha(0) L_0(w), \max_{e_0 \geq e_0^*} (p - w) d - k_0 e_0 - \frac{k_1}{L_1(w)} e_0 L_0(w) \right\}
\]

\( e_0^* = \sqrt{\frac{k_1L_0(w)}{k_0L_1(w)}} \) is optimal if and only if \( L_0(w)L_1(w) \geq \frac{4k_1k_0}{\beta^2} \). Under this condition,

\[
e^* = e_0^* e_1^* = \log \left( \frac{\beta^2 L_0(w) L_1(w)}{k_1 k_0} \right)
\]

Otherwise, \( e_0^* = 0 \) and \( e^* = 0 \). To verify the exact dependence over \( w \), consider \( L_1(w)L_0(w) \). There exists \( w' = \frac{p + c - \alpha(w) - \alpha_0}{2} \) satisfying \( L_1(w) = L_0(w) \) such that when \( w < w' \), \( L_1(w)L_0(w) \) is increasing; when \( w \geq w' \), \( L_1(w)L_0(w) \) is decreasing. Moreover, there exists \( w_L \leq w' \leq w_U \) satisfying \( L_1(w)L_0(w) = \frac{4k_1k_0}{\beta^2} \) such that

\( e_0^* = \sqrt{\frac{k_1L_0(w)}{k_0L_1(w)}} \) when \( w \in [w_L, w_U] \), otherwise \( e_0^* = 0 \).

We now consider the manufacturer’s strategy preference as a function of \( a_i, \gamma \) and \( \beta \). The manufacturer’s optimal profit under the zero effort strategy is

\[
\max_{c \in [w_L, w_U]} \Pi_0^d(w) = (p - w) d - \beta L_0(w)
\]

\[
s.t. \quad \Pi_1^d(w) = (w - c) d - \beta L_1(w) \geq r
\]

Since \( \Pi_0^d(w) \) is decreasing in \( w \) while \( \Pi_1^d(w) \) is increasing, the optimal price \( w^d \) under zero effort strategy satisfies \( \Pi_1^d(w^d) = r \), \( w^d = \frac{r + a_1}{(1 - \gamma)d} + c \). The manufacturer’s optimal profit under non-zero effort strategy is

\[
\max_{w_L \leq w \leq w_U} \Pi_0^e(w) = (p - w) d - 2 \sqrt{\frac{k_1k_0 L_0(w)}{L_1(w)}}
\]

\[
s.t. \quad \Pi_1^e(w) = (w - c) d + \sqrt{\frac{k_0 k_1 L_1(w)}{L_0(w)}} \left( \log \left( \frac{\sqrt{\beta^2 L_0(w) L_1(w)}}{k_1 k_0} \right) - 1 \right) \geq r
\]

118
Since $\Pi_1^e(w)$ is increasing in $w$, there exists $\hat{w}$ so that $\Pi_1^e(w) \geq r$ when $w \geq \hat{w}$. Taking the derivative of $\Pi_0^e(w)$ over $w$, we have

$$\frac{\partial \Pi_0^e(w)}{\partial w} = -d + \gamma d \sqrt{\frac{k_0 k_1 L_0^2}{L_0(w) L_1(w)^3}}$$

Thus $\Pi_0^e(w)$ is concave when $L_1(w) < 3 L_0(w)$, i.e., $L_1(w) < \frac{3}{4} L$. If $L_1(w) > L_0(w)$, i.e., $w > w'$, we have

$$\sqrt{\frac{k_0 k_1 L_0^2}{L_0(w) L_1(w)^3}} < \sqrt{\frac{\beta^2 L_0^2}{4L_1(w)^2}} < \sqrt{\beta^2} = \beta$$

Therefore, $\frac{\partial \Pi_0^e(w)}{\partial w} < 0$ when $w > w'$. Further, $\Pi_0^e(w)$ is concave with respect to $[w_L, w']$ and decreasing when $w > w'$. Denote the unconstrained optimizer of $\Pi_0^e(w)$ as $w_1(a_0)$, which satisfies

$$\gamma^2 k_0 k_1 L_0^2 = L_0(w) L_1(w)^3$$

Let $w^e$ denote the optimal price under non-zero effort strategy, thus the optimal effort is $e^* = \log \sqrt{\frac{\beta^2 L_0(w^e) L_1(w^e)}{k_1 k_0}}$.

We illustrate the analysis for $a_0$. To achieve that, we need to prove the following two parts: (1) when $a_i$ increases, the manufacturer has incentive to choose the non-zero effort strategy over the zero-effort strategy; (2) under the non-zero effort strategy, $e^*$ is increasing in $a_i$. For the first part and for $a_0$, manufacturer’s optimal profit under delegation effort strategy $\Pi_0^d(w)$ has a derivative of $-1$ over $a_0$. For $\Pi_1^e(w)$, taking derivative over $a_0$, we have

$$\frac{\partial \Pi_1^e(w)}{\partial a_0} = - \frac{1}{2} \sqrt{\frac{k_0 L_1(w)}{k_1 L_0(w)^3}} \left( \log \left( \sqrt{\frac{k_0 k_1}{\beta^2 L_0(w) L_1(w)}} \right) - 2 \right)$$
For \( w \in [w_L, w_U] \) with \( \frac{k_0k_1}{\beta^2L_0(w)L_1(w)} \leq \frac{1}{4} \), \( \frac{\partial \Pi^c_1(w)}{\partial a_0} > 0 \), \( \Pi^c_1(w) \) is increasing in \( a_0 \) on \([w_L, w_U] \) with the range \([w_L, w_U] \) is expanding over \( a_0 \). Lastly, taking the derivative of \( \Pi^c_0(w) \) with respect to \( a_0 \), we have

\[
\frac{\partial \Pi^c_0(w)}{\partial a_0} = -\sqrt{\frac{k_0k_1}{L_0(w)L_1(w)}} > -\frac{1}{2}
\]

Thus the manufacturer’s optimal profit under the control strategy decreases slower in \( a_0 \), compared with that under delegation strategy. Therefore, for \( a_0 \leq a'_0 \), if control strategy is optimal under \( a_0 \), it is also optimal under \( a'_0 \). For the second part, to verify the monotonicity of \( e^* \) over \( a_i \), it is sufficient to verify that of \( L_0(w^c)L_1(w^c) \) over \( a_i \). Since \( \Pi^c_1(w) \) is increasing in \( w \), denote the minimal \( w \) that satisfies the Tier 1 supplier reservation profit constraint as \( \hat{w}(a_0) \) and \( \hat{w}(a_0) \) is decreasing in \( a_0 \). Thus the optimal price \( w^c = \max \{\hat{w}(a_0) , w_1(a_0)\} \). When \( w^c = w_1(a_0) \), if \( w_1(a_0) \) increases, \( L_0(w^c)L_1(w^c) \) increases given that \( w_1(a_0) < w_c(a_0) \). If \( w_1(a_0) \) decreases, then \( L_1(w_1(a_0)) \) decreases. Therefore, \( L_0(w^c)L_1(w^c) \) must increase since the left hand side of the equation is increasing. When \( \hat{w}(a_0) \geq w_1(a_0) \), \( w^c = \hat{w}(a_0) \). Since \( \hat{w}(a_0) \) is decreasing in \( a_0 \), \( L_0(w^c)L_1(w^c) \) increases. Thus \( e^* \) is increasing. The analysis applies similarly for \( a_1, \gamma, \) and \( \beta \). \( \square \)

**Proof of Lemma 11**

*Proof.* Based on the assumption that \( b > \max \{c, w\} \), it is always optimal to deploy the resources to satisfy demand. Since the total demand \( D \) is known, if \( m > D \), there is no need to produce from reactive capacity, i.e. \( p^*_\mu(t) = 0 \). Otherwise, it is optimal to produce at full rate \( p^*_\mu(t) = \mu \), targeting \( m + \mu t \) reaches total demand \( D \).

When shortage firstly occur at a time \( t' \), there are three cases.

**Case 1:** If \( x(s) \) is increasing or stationary, \( x(t') \geq \mu \) (otherwise, \( x(s) < \mu \) for \( 0 \leq s \leq t' \), no shortage at time \( t' \)). Given there is no safety stock leftover and
$x(s) \geq \mu$ for $s \geq t'$, it is optimal to produce to satisfy as much as possible, i.e.

$p^*_\mu(s) = \mu = \min \{x(s), \mu\}$ for $s \geq t'$.

**Case 2:** If $x(s)$ is decreasing, $x(t') \geq \mu$. Otherwise, if $x(t') < \mu$, we have

$$m + \mu t < \int_0^{t'} x(s) \, ds.$$  
However, for $t' - \delta t$,

$$m + \mu(t' - \delta t) = m + \mu t' - \mu \delta t$$

$$\leq \int_0^{t'} x(s) \, ds - \mu \delta t < \int_0^{t' - \delta t} x(s) \, ds$$

This contradict the definition of $t'$. Given $x(t') \geq \mu$ and $x(s)$ is decreasing, if there exist $t' \leq t'' < T(\omega)$ so that $x(t'') = \mu$, it is optimal to produce at rate $p^*_\mu(s) = \mu = \min \{x(s), \mu\}$ before $t''$, to produce at demand $p^*_\mu(s) = x(s) = \min \{x(s), \mu\}$ after $t''$. If $t''$ does not exist, i.e. $x(T(\omega)) \geq \mu$, it is optimal to produce at rate $p^*_\mu(s) = \mu = \min \{x(s), \mu\}$ after $t'$.

**Case 3:** If $x(s)$ is first increasing then decreasing, similarly we can show that $x(t') \geq \mu$. Using the same logic as in Case 2, we can prove that $p^*_\mu(s) = \mu = \min \{x(s), \mu\}$.

In summary, it is optimal to produce at full rate before shortage occur. Once shortage occurs, the manufacturer produces at rate $p^*_\mu(s) = \min \{x(s), \mu\}$. The optimal deployment policy is immediate if $c \leq w$; in this case, it is optimal to deploy the cheaper and available stock first.

**Proof of Lemma 12.**

**Proof.** This result can be directly derived from Definition 1. Here we only show the justification when $x(t)$ is increasing, the reasoning for the other cases is similar. (1) If $x(T) \leq \mu$, stockout can never happen, the service level is equivalent to demand satisfaction at any point $t$. (2) If $x(0) \geq \mu$, then $m + \mu t - \int_0^t x(s) \, ds$ is decreasing.
in $t$, therefore $P\left\{ \int_0^t X(s) \, ds \leq m + \mu t, \forall 0 \leq t \leq T \right\} = P\left\{ \int_0^T X(s) \, ds \leq m + \mu T \right\}$.

Define $T'(\mu) = T$. (3) Else $x(0) < \mu < x(T)$, then $m + \mu t - \int_0^t x(s) \, ds$ is increasing before the intersection point and decreasing after the intersection point. Since $m + \mu t - \int_0^t x(s) \, ds \bigg|_{t=0} = m > 0$, stock-out can only happen after the intersection point.

Therefore, the possible stock-out point with lowest $m + \mu t - \int_0^t x(s) \, ds$ is $T$. In summary, $T'(\mu) = T$.

**Proof of Lemma 13.**

**Proof.** Recall $B'(\mu) = D\left(T'(\mu)\right) - \mu T'(\mu)$, we need to prove that $B'(\mu)$ is convex with respect to $\mu$ for each $\omega$.

**Case 1:** When $x(t, \omega)$ is increasing or stationary, then $T'(\mu) = T$. Then $D\left(T'(\mu)\right) = D$, independent of $\mu$, $m$. $B'(\mu) = D - \mu T$ is convex with respect to $\mu$.

**Case 2:** When $x(t, \omega)$ is first increasing then decreasing or decreasing

$$B'(\mu) = \begin{cases} D - \mu T, & \mu < x(T) \\ \int_0^T x(s) \, ds - \mu T'(\mu), & x(T) \leq \mu \leq x(t_p) \\ 0, & \mu > x(t_p) \end{cases}$$

It is easy to verify that $B'(\mu)$ is decreasing and continuous in $\mu$ and $B'(\mu)$ is convex when $\mu < x(T)$ and $\mu > x(t_p)$.

In order to prove that $B'(\mu)$ is convex in general, we need to show that $B'(\mu)$ is convex with respect to $\mu$ on interval $x(T) \leq \mu \leq x(t_p)$. If assuming that $T'$ is differentiable with respect to $\mu$, we have

$$0 > \frac{\partial \int_0^T x(t) \, dt - \mu T'(\mu)}{\partial \mu} = -T' > -T$$

$$\frac{\partial^2 \int_0^T x(t) \, dt - \mu T'(\mu)}{\partial^2 \mu} = \frac{\partial - T'}{\partial \mu} > 0$$
Thus $B' (\mu)$ is convex. For the case when $T' (\mu)$ is not differentiable, the results can be similarly derived using differences instead of derivatives.

Given $B' (\mu)$ is convex with respect to $\mu$ for each $\omega$ and convexity is preserved under operator $\max \{ x, 0 \}$ and expectation, we have $E [B' (\mu)]^+$ and $E [B' (\mu) - m]^+$ are also jointly convex in $(m, \mu)$. $c - w > 0$, $b > c$, therefore $G^p(m, \mu)$ is jointly convex in $(m, \mu)$. Next, we need to show the existence of the optimal solution. We have $G^p(0, 0) = \frac{bE[D]}{\nu + \tau}$, $m \to +\infty G^p(m, \mu) \to +\infty$ and $\mu \to +\infty G^p(m, \mu) \to +\infty$. There exist $(m', \mu')$ such that when $(m, \mu) > (m', \mu')$, $G^p(m, \mu) \geq \frac{bE[D]}{\nu + \tau}$. Thus the optimization is equivalent to minimize $G^p(m, \mu)$ over feasible set $\{ 0 \leq m \leq m', 0 \leq \mu \leq \mu' \}$. According to Weierstrass Theorem, the optimal solution is guaranteed to exist. At last, since the constraints $m \geq 0$ and $\mu \geq 0$ are both concave and satisfy Slater’s condition, according to the theorem of Kuhn and Tucker under Convexity, we have $(m, \mu) \geq 0$ is optimal if and only if it satisfies the Kuhn-Tucker first-order condition.

Proof of Proposition 14

Proof. According to the Khun Tucker condition, $(m, \mu) \geq 0$ is optimal if and only if there exists $\lambda_1, \lambda_2 \geq 0$ so that

$$h \nu = (w - c) P \{ D > m \} + (b - w) P \{ B' (\mu) > m \} + \lambda_1$$

$$r \nu = (b - w) E \left[ T' (\mu) | B' (\mu) > m \right] P \{ B' (\mu) > m \} + \lambda_2$$

$$\lambda_1 m = 0, \lambda_2 \mu = 0$$

From this condition, we can derive the optimal solution $(0, 0)$, $(m_1, 0)$, $(0, \mu_1)$ and $(m_2, \mu_2)$. Namely,

Case 1: $(0, 0)$ is optimal if and only if $h \nu \geq (w - c) + (b - w)$ and $r \nu \geq (b - w) \tau$, i.e., $h \geq \frac{b - c}{\nu}$ and $r \geq \frac{(b - w) \tau}{\nu}$.  

123
Case 2: $m_1$ exists only if $h < \frac{b-c}{\nu}$. $(m_1, 0)$ is optimal if and only if $rv - (b - w) E [T|D (T) > m_1] P \{D (T) > m_1\} \geq 0$. Thus $(m_1, 0)$ is optimal if and only if $h < \frac{b-c}{\nu}$ and $r \geq r^* (h)$.

Case 3: $\mu_1$ exists only if $r < \frac{(b-w)\tau}{\nu}$. $(0, \mu_1)$ is optimal if and only if $h \nu - (w - c) - (b - w) P \{B' (\mu_1)\} \geq 0$. Thus $(0, \mu_1)$ is optimal if and only if $r < \frac{(b-w)\tau}{\nu}$ and $h \geq h^* (r)$.

Case 4: For $r < r^* (h)$ and $h < h^* (r)$, the optimal solution $(m^*, u^*)$ is none of the above form. Thus, $m^* > 0$ and $\mu^* > 0$. As a result, $(m^*, \mu^*) = (m_2, \mu_2)$.

Summarizing these results, we have Proposition 1. For the monotonicity, we only illustrate the proof for $r^* (h)$. When $h < \frac{b-c}{\nu}$, if $h$ increases, $m_1 (h)$ decreases, thus condition $D > m_1 (h)$ holds on larger set of $\omega$, thus $E [T \mid D > m_1 (h)] P \{D > m_1 (h)\}$ increases. Further, $r^* (h) \leq \frac{(b-w)\tau}{\nu}$. Therefore $r^* (h)$ increases globally.

Proof of Proposition 15

Proof. When $b$ increases, $(0, 0)$ is less likely to be optimal and the service level is more likely to be non-zero. When the service level is non-zero, consider the problem case by case:

Case 1: If the optimal solution is of form $(m_1, 0), P \{D (T) > m_1\} = \frac{hv}{b-c}$, the service level $\alpha^* = 1 - \frac{hv}{b-c}$ is increasing in $b$.

Case 2: If the optimal solution is of form $(0, \mu_1), E [T' (\mu) \mid B' (\mu) > 0] P \{B' (\mu) > 0\} = \frac{r\nu}{b-w}$. Since $B' (\mu)$ and $T' (\mu)$ is decreasing in $\mu$, $E [T' (\mu) \mid B' (\mu) > 0] P \{B' (\mu) > 0\}$ is also decreasing in $\mu$. Therefore, $\mu_1$ is increasing in $b$, therefore $\alpha^* = 1 - P \{B' (\mu_1) > 0\}$ is increasing in $b$.

Case 3: If joint deployment is optimal,

$$(b - w) P \{B' (\mu_2) > m_2\} + (w - c) P \{D (T) > m_2\} = hv$$

$$(b - w) E [T' (\mu_2) \mid B' (\mu_2) > m_2] P \{B' (\mu_2) > m_2\} = rv$$

124
we know that at least one of \( m_2 \) and \( \mu_2 \) increases in \( b \). If \( m_2 \) decreases, \( P \{ B' (\mu_2) > m_2 \} \) is decreasing. If \( m_2 \) increases and \( P \{ B' (\mu_2) > m_2 \} \) is non-decreasing, then \( \mu_2 \) decreases. As will be shown in Proposition 16, this will lead to the left side of the second equation increasing, which contradicts the second equation. Thus, \( P \{ B' (\mu_2) > m_2 \} \) can only be decreasing, so \( \alpha^* = 1 - P \{ B' (\mu_2) > m_2 \} \) is increasing.

Since in each region, \( \alpha^* \) is increasing in \( b \). Moreover, the optimal policy is continuous on the boundaries. Thus, \( \alpha^* \) is increasing in \( b \) globally. When \( b = 0 \), \((m^*, \mu^*) = (0, 0)\), thus \( \alpha^* = 0 \). When \( b \to +\infty \), \( m^* \) or \( \mu^* \to +\infty \), thus \( \alpha^* = 1 \). Therefore, for any chosen desired service level \( \alpha^* \), there exists an \( b \) such that the optimal service level achieved by the optimal solution under the unit penalty cost constraint is \( \alpha^* \). Namely, the induced unit penalty constraint is equivalent to the service level constraint.  

**Proof of Proposition 16**

*Proof.* We need to show the monotonicity of the optimal solution \((m^*, \mu^*)\) in the cost parameter \( h, c, w, r, b \). Due to space limit, we only prove the results for \( h \) and \( b \), the proof of the other results follows the same logic.

When the optimal solution is of the form \((m_1, 0)\), \( m_1 \) is decreasing in \( h \). When the optimal solution is of the form \((0, \mu_1)\), the optimal \( m^* = 0 \) is non-increasing in \( h \). When the optimal solution is of the form \((m_2, \mu_2)\), we denote \((m_2 (h), \mu_2 (h))\) to imply the dependence on \( h \). As we have shown in Lemma 13, \( T' (\mu, \omega) \) and \( B' (\mu, \omega) \) are decreasing functions in \( \mathbb{R} \) for each \( \omega \). Consider \( h_2 = h_1 + \delta h \) and compare \((m_2 (h), \mu_2 (h))\) under two cases. From the first equation, we know that at least one of \( m_2 (h_2) < m_2 (h_1) \) and \( \mu_2 (h_2) < \mu_2 (h_1) \) holds in order to guarantee the left side of the first equation be increasing in \( h \). From the second equation, we know that at least one of \( m_2 (h_2) > m_2 (h_1) \) and \( \mu_2 (h_2) > \mu_2 (h_1) \) holds in order to guarantee the right side of the second equation to be constant. If \( m_2 (h_2) > m_2 (h_1) \) and
\[ \mu_2(h_2) < \mu_2(h_1), \ T'(\mu_2(h_2), \omega) \geq T'(\mu_2(h_1), \omega), \ B'(\mu_2(h_2), \omega) \geq A(\mu_2(h_1), \omega). \]

Let \( \Delta_i = \{ \omega : B'(\mu_2(h_i), \omega) > m_2(h_i) \} \), from the first equation we have \( P\{\Delta_2\} > P\{\Delta_1\} \) and thus \( P\{\Delta_2 \setminus \Delta_1\} > P\{\Delta_1 \setminus \Delta_2\} \).

\[
\int_{\Delta_2} T'(\mu_2(h_2), \omega) \, d\mathbb{P}(\omega) - \int_{\Delta_1} T'(\mu_2(h_1), \omega) \, d\mathbb{P}(\omega) \\
= \int_{\Delta_2 \cap \Delta_1} T'(\mu_2(h_2), \omega) - T'(\mu_2(h_1), \omega) \, d\mathbb{P}(\omega) \\
+ \int_{\Delta_2 \setminus \Delta_1} T'(\mu_2(h_2), \omega) \, d\mathbb{P}(\omega) - \int_{\Delta_1 \setminus \Delta_2} T'(\mu_2(h_1), \omega) \, d\mathbb{P}(\omega)
\]

For \( \omega \in \Delta_2 \setminus \Delta_1 \), \( B'(\mu_2(h_2), \omega) > m_2(h_2) \) while \( B'(\mu_2(h_1), \omega) < m_2(h_1) \), therefore \( B'(\mu_2(h_2), \omega) - B'(\mu_2(h_1), \omega) > m_2(h_2) - m_2(h_1) \). For \( \omega \in \Delta_1 \setminus \Delta_2 \), \( B'(\mu_2(h_2), \omega) \leq m_2(h_2) \) while \( B'(\mu_2(h_1), \omega) > m_2(h_1), \) therefore \( B'(\mu_2(h_2), \omega) - B'(\mu_2(h_1), \omega) < m_2(h_2) - m_2(h_1) \). Since \( B'(\mu_2(h_2), \omega) - B'(\mu_2(h_1), \omega) = \int_{\mu_2(h_1)}^{\mu_2(h_2)} T'(\omega, \omega) \, ds \), if we consider the extreme case that \( 0 < \delta h \) infinitely small, the above result means that for any \( \omega_1 \in \Delta_2 \setminus \Delta_1 \) and \( \omega_2 \in \Delta_1 \setminus \Delta_2 \), \( T'(\mu_2(h_2), \omega), \omega_1) > T'(\mu_2(h_2), \omega_2) \). Therefore

\[
\int_{\Delta_2 \setminus \Delta_1} T'(\mu_2(h_2), \omega) \, d\mathbb{P}(\omega) - \int_{\Delta_1 \setminus \Delta_2} T'(\mu_2(h_1), \omega) \, d\mathbb{P}(\omega) \\
\geq \int_{\Delta_2 \setminus \Delta_1} T'(\mu_2(h_2), \omega) \, d\mathbb{P}(\omega) - \int_{\Delta_1 \setminus \Delta_2} T'(\mu_2(h_2), \omega) \, d\mathbb{P}(\omega) > 0 \\
\implies \int_{\Delta_2} T'(\mu_2(h_2), \omega) \, d\mathbb{P}(\omega) - \int_{\Delta_1} T'(\mu_2(h_1), \omega) \, d\mathbb{P}(\omega) > 0
\]

This contradicts the second equation. Therefore \( m_2(h_2) < m_2(h_1) \) and \( \mu_2(h_2) > \mu_2(h_1) \). In each region, \( m^* \) is decreasing in \( h \) and \( \mu^* \) is increasing in \( h \). Since the optimal solution region is continuous, we can conclude \( m^* \) is decreasing in \( h \) and \( \mu^* \) is increasing in \( h \). Further,

\[
\gamma_m(h) = \frac{m^*(h)}{m^*(h) + E[D - B'(\mu^*(h))]}\]

126
Since $B' (\mu)$ is decreasing in $\mu$ stochastically, $E \left[ D - B' (\mu^* (h)) \right]$ increases in $h$. Therefore $\gamma_m (h)$ is decreasing.

For the proof for the penalty cost $b$, we know that at least one of $m_2$ and $\mu_2$ increases in $b$. Consider $b_2 = b_1 + \delta b$, if $\mu_2 (b_2) < \mu_2 (b_1)$, $m_2 (b_2) > m_2 (b_1)$. From the first equation, we have

$$(b_2 - w) P \left\{ B' (\mu_2 (b_2)) > m_2 (b_2) \right\} \geq (b_1 - w) P \left\{ B' (\mu_2 (b_1)) > m_2 (b_1) \right\}$$

In addition, for any $\omega_1 \in \Delta_2 \setminus \Delta_1$ and $\omega_2 \in \Delta_1 \setminus \Delta_2$, $T' (\mu_2 (h_2), \omega_1) > T' (\mu_2 (h_2), \omega_2)$. Using a similar method as above, we can show $(b-w) E \left[ T' \mid B' (\mu) > m \right] P \left\{ B' (\mu) > m \right\}$ is increasing in $b$, a contradiction. Therefore $\mu_2 (b_2) \geq \mu_2 (b_1)$, i.e., $\mu^*$ is increasing in $b$.

\begin{proof}
Proof of Proposition 17

Proof. From the property of the Lorenz order, there exists $k > 1$ so that $T_2 \sim k T_1$. If $T_1$ increases to $T_2 = k \cdot T_1$, $X_2 (t, \omega) = X_1 \left( \frac{t}{k}, \omega \right)$, we can show that $m_1 (T_2) \leq k \cdot m_1 (T_1)$, $\mu_1 (T_2) \geq \mu_1 (T_1)$. $T' (\mu, T_2) = k T' (\mu, T_1)$, $D (T' (\mu, T_2)) = k D (T' (\mu, T_1))$ and $B' (\mu, T_2) = k B' (\mu, T_1)$. When joint strategy is optimal, the optimal solution $(m_2, \mu_2)$ satisfies

$$(b-w) P \left\{ B' (\mu, T_1) > \frac{m}{k} \right\} + (w-c) P \left\{ D (T_1) > \frac{m}{k} \right\} = h \nu$$

$$E \left[ T' (\mu, T_1) \mid B' (\mu, T_1) > \frac{m}{k} \right] P \left\{ B' (\mu, T_1) > \frac{m}{k} \right\} = \frac{r \nu}{k}.$$ 

It is easy to check that under $(km_2 (T_1), \mu_2 (T_1))$, the first equation holds, while the left side of the second equation is greater than the right side. Since $E \left[ T' (\mu) \mid B' (\mu) > \frac{m}{k} \right] P \left\{ B' (\mu) > \frac{m}{k} \right\}$ is decreasing in $\mu$ and $m$, at least one of $m_2 (T_2) > km_2 (T_1)$ or $\mu_2 (T_2) > \mu_2 (T_1)$ holds. From the first equation, if $m_2 (T_2) > km_2 (T_1)$, then $\mu_2 (T_2) < \mu_2 (T_1)$, $P \left\{ B' (\mu, T_1) > \frac{m_2 (T_2)}{k} \right\} > P \left\{ B' (\mu, T_1) > m_2 (T_1) \right\}$.

127
From the proof of Proposition 3, we know that if \( \frac{m_2(T_2)}{k} > m_2(T_1) \) and \( \mu_2(T_2) < \mu_2(T_1) \), then the left side of the second equation must increase, a contradiction. Therefore, \( m_2(T_2) \leq k m_2(T_1) \) and \( \mu_2(T_2) > \mu_2(T_1) \). Further, based on the formula of \( h^* (r) \) and \( r^* (h) \), we have \( P \{ B' (\mu_1) \} \) decreases and \( h^* (r) \) decreases in \( k \). At the same time, for each given \( \omega \), \( D(T_2) - m_1(h, T_2) \geq k (D(T_2) - m_1(h, T_1)) \) and \( r^* (h) \) increases in \( k \). Thus \((m_1, 0)\) has smaller region to be optimal and \((0, \mu_1)\) has larger region to be optimal. Therefore, \( m^* (T_2) \leq k * m^* (T_1) \) and \( \mu^* (T_2) \geq \mu^* (T_1) \).

Moreover, \( E[D(T_2) - B' (\mu, T_2)] = k E[D(T_1) - B' (\mu, T_1)] \), therefore

\[
\gamma_m(T_2) = \frac{m^*(T_2)}{m^*(T_2) + k E[D(T_1) - B' (\mu^* (T_2), T_1)]} \\
\leq \frac{m^*(T_1)}{m^*(T_1) + E[D(T_1) - B' (\mu^* (T_2), T_1)]} \\
\leq \frac{m^*(T_1)}{m^*(T_1) + E[D(T_1) - B' (\mu^* (T_1), T_1)]} = \gamma_m(T_1)
\]

The strategy preference stays the same.

**Proof of Proposition 18**

**Proof.** The analysis for the rectangular and triangular surge demands is similar. Thus due to space limit, we only give the proof for rectangular surge demand. Under rectangular disaster demand \( T' (\mu) = T \) and \( D(T) - B' (\mu, T) = \mu T \). When \( T \) increases according to the above convex order, \( P \{ X > \mu_1 \} = \frac{rv}{(b-w)r} \) is constant. It is easy to check that \( E [T | XT > m_1] P \{ XT > m_1 \} \) is increasing in \( k \) given \( P \{ XT > m_1 \} = \frac{hw}{b-c} \). Moreover, \( m'_1 (k) k - m_1 < 0 \) and \( \mu_1 (k) \) stays constant. Therefore when \( k \) increases, \( h^* (r) \) stays constant while \( r^* (h) \) decreases. From this, we know the region of optimal \((m_1, 0)\) is shrinking, the region of optimal \((0, \mu_1)\) is constant while the region of optimal \((m_2, \mu_2)\) is expanding.
When joint policy is optimal,

\[(b - w) \int_{k^{-1}\tau}^{+\infty} \tilde{F} \left( \frac{m_2}{k(y - \tau) + \tau} + \mu_2 \right) dG(y) + \]

\[(w - c) \int_{k^{-1}\tau}^{+\infty} \tilde{F} \left( \frac{m_2}{k(y - \tau) + \tau} \right) dG(y) = h\nu\]

\[(b - w) \int_{k^{-1}\tau}^{+\infty} (k(y - \tau) + \tau) \tilde{F} \left( \frac{m_2}{k(y - \tau) + \tau} + \mu_2 \right) dG(y) = r\nu\]

Take derivative on both sides and simplify, we have

\[(b - w) \int_{0}^{+\infty} f \left( \frac{m_2}{t} + \mu_2 \right) \left( \frac{m_2'(k) - m_2}{t} + \frac{m_2\tau}{t^2} + \mu_2'(k) \right) dG \left( \frac{t - \tau}{k} + \tau \right)\]

\[+ (w - c) \int_{0}^{+\infty} f \left( \frac{m_2}{t} \right) \left( \frac{m_2'(k) - m_2}{t} + \frac{m_2\tau}{t^2} \right) dG \left( \frac{t - \tau}{k} + \tau \right) = 0\]

\[(b - w) \int_{0}^{+\infty} tf \left( \frac{m_2}{t} + \mu_2 \right) \left( \frac{m_2'(k) - m_2}{t} + \frac{m_2\tau}{t^2} + \mu_2'(k) \right) dG \left( \frac{t - \tau}{k} + \tau \right)\]

\[- (b - w) \int_{0}^{+\infty} \frac{t - \tau}{k} \tilde{F} \left( \frac{m_2}{t} + \mu_2 \right) dG \left( \frac{t - \tau}{k} + \tau \right) = 0\]

From equation 1 at least one of \(m_2'(k) - \frac{m_2}{k}\) and \(\mu_2'(k) < 0\). If \(m_2'(k) - \frac{m_2}{k} > 0\), then \(\mu_2'(k) < 0\). Therefore, \(\int_{0}^{+\infty} f \left( \frac{m_2}{t} + \mu_2 \right) \left( \frac{m_2'(k) - m_2}{t} + \frac{m_2\tau}{t^2} + \mu_2'(k) \right) dG \left( \frac{t - \tau}{k} + \tau \right) < 0\). As a result, we can show that

\(\int_{0}^{+\infty} tf \left( \frac{m_2}{t} + \mu_2 \right) \left( \frac{m_2'(k) - m_2}{t} + \frac{m_2\tau}{t^2} + \mu_2'(k) \right) dG \left( \frac{t - \tau}{k} + \tau \right) < 0\). Given that \(T_1\) is symmetric with respect to \(\tau\), we have \(\int_{0}^{+\infty} \frac{t - \tau}{k} \tilde{F} \left( \frac{m_2}{t} + \mu_2 \right) dG \left( \frac{t - \tau}{k} + \tau \right) > 0\), a contradiction. Therefore, \(m_2'(k) - \frac{m_2}{k} < 0\). By revising the formula using \(dF(x)\) as the integral, we can similarly show that \(\mu_2'(k) > \frac{m_2}{k}\). Therefore \(\mu_2\) is increasing in \(k\).
Furthermore,
\[
\gamma_m(T) = \frac{m_2}{m_2 + E[D(T) - B'(\mu_2(T), T)]} \\
= \frac{m_2}{m_2 + \mu_2(T)\tau}
\]
Thus \(\gamma_m(T)\) is decreasing when \(T\) changes from \(T_1\) to \(T_2\).

**Proof of Proposition 19**

*Proof.* For Lorenz order of \(X_1(t)\) and \(X_2(t)\), there exists \(k \geq 1\) so that \(X_2(t, \omega) \sim kX_1(t, \omega)\). When surge demand increases from \(X_1(t, \omega)\) to \(X_2(t, \omega)\), \(T'(\mu, X_2) = T'\left(\frac{\mu}{k}, X_1\right), \ D\left(T'(\mu, X_2), X_2\right) = k * D\left(T'\left(\frac{\mu}{k}, X_1\right), X_1\right)\). Thus
\[
B'(\mu, X_2) = D\left(T'(\mu, X_2), X_2\right) - \mu T'(\mu, X_2) \\
= k * D\left(T'\left(\frac{\mu}{k}, X_1\right), X_1\right) - k * \frac{\mu}{k} T'\left(\frac{\mu}{k}, X_1\right) \\
= kB'\left(\frac{\mu}{k}, X_1\right)
\]
Substituting this fact into the optimal policy equations, we can get the optimal policy as \((km^*(X_1), k\mu^*(X_1))\). Therefore,
\[
\gamma_m(X_2) = \frac{m^*(X_2)}{m^*(X_2) + E[D(X_2) - B'(\mu^*(X_2), X_2)]} \\
= \frac{km^*(X_1)}{km^*(X_1) + E[kD(X_1) - kB'(\mu^*(X_1), X_1)]} \\
= \frac{m^*(X_1)}{m^*(X_1) + E[D(X_1) - B'(\mu^*(X_1), X_1)]} \\
= \gamma_m(X_1)
\]

**Proof of Proposition 20**
Proof. Based on results from Proposition 17 and 19, when change $X_1 \rightarrow X_2$ and $T$ stays constant, we have $m^*(X_2, T_1) = \frac{1}{k} m^*(X_1, T_1)$, $\mu^*(X_2, T_1) = \frac{1}{k} \mu^*(X_1, T_1)$, and $\gamma_m(X_2, T_1) = \gamma_m(X_1, T_1)$. Then when change $T_1 \rightarrow T_2$ and $X$ stays constant, $m^*(X_2, T_2) \leq km^*(X_2, T_1)$, $\mu^*(X_2, T_2) \geq \mu^*(X_2, T_1)$ and $\gamma_m(X_2, T_2) \leq \gamma_m(X_2, T_1)$. Therefore, $m^*(X_2, T_2) \leq m^*(X_1, T_1)$ and $\mu^*(X_2, T_2) \geq \frac{\mu^*(X_1, T_1)}{k}$. Obviously, $\gamma_m(X_2, T_2) \leq \gamma_m(X_1, T_1)$.

\[ \square \]

Proof of Proposition 21

Proof. Under the demand group $X'(t, \beta, \omega)$, we have $T'(\mu, \beta) = \beta T + (1 - \beta) T'(\mu, 0)$. Here $T'(\mu, 0)$ is critical time when $\beta = 0$ and $X'(t, 0)$ is decreasing.

\[
B'(\mu, \beta) = \int_0^{T'(\mu, \beta)} X'(t, \beta) \, dt - \mu T'(\mu, \beta)
\]

\[
= \int_0^{\beta T} X'(t, \beta) \, dt + \int_{\beta T}^{T'(\mu, \beta)} X'(t, \beta) \, dt - \mu \left( \beta T + (1 - \beta) T'(\mu, 0) \right)
\]

\[
= \int_0^{\beta T} X(t) \left( \frac{t}{\beta} \right) \, dt - \mu \beta T + \int_{\beta T}^{T'(\mu, \beta)} X \left( \frac{t - \beta T}{1 - \beta} \right) \, dt - \mu (1 - \beta) T'(\mu, 0)
\]

\[
= \beta (D(T) - \mu T) + (1 - \beta) \left( \int_0^{T'(\mu, 0)} X(t) - \mu dt \right)
\]

\[
= \int_0^{T'(\mu, 0)} X(t) - \mu dt + \beta \int_{T'(\mu, 0)}^{T} X(t) - \mu dt
\]

When $\beta$ increases, $m_1$ stays the same, $\mu_1(\beta)$ satisfies

\[
E \left[ \beta T + (1 - \beta) T'(\mu, 0) \left| B'(\mu, \beta) > 0 \right. \right] P \left\{ B'(\mu, \beta) > 0 \right\} = \frac{r\nu}{b - w}
\]

Since for every $t \in [T'(\mu, 0), T]$, $X(t) \leq \mu$ according to the definition of $T'(\mu, 0)$, $B'(\mu, \beta)$ is decreasing in $\beta$ and $\mu$. $\beta T + (1 - \beta) T'(\mu, 0)$ is increasing in $\beta$, but
decreasing in $\mu$. If $\beta_1 < \beta_2$ and $\mu_1(\beta_1) \leq \mu_1(\beta_2)$, $P \{B'(\mu_1(\beta), \beta) > 0\}$ is decreasing in $\beta$. If $\mu_1(\beta_1) > \mu_1(\beta_2)$, $\beta T + (1 - \beta) T'(\mu_1(\beta), 0)$ is increasing in $\beta$. If $P \{B'(\mu_1(\beta), \beta) > 0\}$ is increasing, define $\Delta_{(\beta, \mu)} = \{\omega : B'(\mu, \beta) > 0\}$ and then $P \{B'(\mu, \beta) > 0\} = \int_{\Delta_{(\beta, \mu)}} d\mathbb{P}(\omega)$. If the assumption is true,

$$
\int_{\Delta_{(\beta_1, \mu_1(\beta_1))}} d\mathbb{P}(\omega) < \int_{\Delta_{(\beta_2, \mu_1(\beta_2))}} d\mathbb{P}(\omega)
$$

Similarly, the left side of the second equation can be expressed as $\int_{\Delta_{(\beta, \mu)}} \beta T + (1 - \beta) T'(\mu, 0) d\mathbb{P}(\omega)$. For $\omega \in \Delta_{(\beta_1, \mu_1(\beta_1))} \cap \Delta_{(\beta_2, \mu_1(\beta_2))}$, $\beta T + (1 - \beta) T'(\mu, 0)$ is increasing in $\beta$. Furthermore, given $\omega$, $\frac{\partial B'(\mu, \beta)}{\partial \mu} = -\beta T - (1 - \beta) T'(\mu, 0)$. Therefore, adopting a similar method as in Proposition 2, the left side of the above equation is increasing, a contradiction. As a result, $P \{B'(\mu_1(\beta), \beta) > 0\}$ is always decreasing. Thus, $r^*(h)$ stays constant while $h^*(r)$ decreases, so $(m_2, \mu_2)$ has smaller region to be optimal while $(0, \mu_1)$ has larger region to be optimal.

When joint policy is optimal,

$$(b - w) P \{B'(\mu, \beta) > m\} + (w - c) P \{D(T) > m\} = h\nu$$

$$(b - w) E \left[ \beta T + (1 - \beta) T'(\mu, 0) \middle| B'(\mu, \beta) > m \right] P \{B'(\mu, \beta) > m\} = r\nu$$

From equation 1, at least one of $m$ and $\mu$ should be decreasing in $\beta$. If $m_2(\beta_2) > m_2(\beta_1)$, then $\mu_2(\beta_2) < \mu_2(\beta_1)$ and $P \{B'(\mu_2(\beta), \beta) > m_2(\beta)\}$ is increasing. Using the similar method as previous, we can show that the left-side of the second equation is increasing, a contradiction. Therefore, $m^*(\beta_2) \leq m^*(\beta_1)$, and the service level $\alpha^* = 1 - P \{B'(\mu_2(\beta), \beta) > m_2(\beta)\}$ is decreasing.

$$\gamma_m(\beta) = \frac{m_2(\beta)}{m_2(\beta) + E[D(T) - B'(\mu_2(\beta), \beta)]}$$
It is easy to confirm that $E[D(T)]$ stays the same. For $E[B'(\mu_2(\beta), \beta)]$ is decreasing in $\beta$ and $\mu_2(\beta)$. When $\mu_2(\beta)$ increases, $E[B'(\mu_2(\beta), \beta)]$ decreases and $\gamma_m(\beta)$ decreases. When $\mu_2(\beta)$ decreases, since $\frac{\partial B'(\mu, \beta)}{\partial \mu} = -\beta T - (1 - \beta) T'(\mu, 0)$, and $-\beta T - (1 - \beta) T'(\mu, 0)$ decreases in $\beta$ and increases in $\mu$. Therefore, $\frac{\partial B'(\mu, \beta)}{\partial \mu}$ decreases. Therefore we have $E[B'(\mu_2(\beta), \beta)]$ decreases and still $\gamma_m(\beta)$ decreases. Therefore $\gamma_m(\beta)$ decreases generally.

Proof of Proposition 22

Proof. For the specific demand, we have $T'(\mu, \rho) = T'(\frac{\mu - \rho X}{(1 - \rho)}, 0)$, here $T'(\mu, 0)$ corresponds the case when there is no rectangular demand base. Using a similar method as in the proof of Proposition 21, we can show that $m_2$ is decreasing and $\alpha^*$ is increasing in $\rho$. The details are omitted here.

Proof of Proposition 23

Proof. For rectangular surge demand,

$$h^*(r) = \begin{cases} \frac{w-c}{\nu} + \frac{r}{\tau}, & r < \frac{(b-w)\tau}{\nu} \\ \frac{b-c}{\nu}, & \text{otherwise} \end{cases}$$

$$r^*(h) = \begin{cases} \frac{h b-c}{b-c} E[T | D > m_1(h)] & h < \frac{b-c}{\nu} \\ \frac{(b-w)\tau}{\nu}, & \text{otherwise} \end{cases}$$

Here, we only illustrate the results when $\kappa \leq 1$, i.e., $\frac{r}{h} \leq \frac{(b-w)\tau}{b-c}$. When $\lambda \to 0$, $\nu \to +\infty$. Thus $r \geq \frac{(b-w)\tau}{\nu}$ and $h^*(r) = \frac{b-c}{\nu} \to 0$, $h \geq h^*(r)$. So $m^* = 0$ and $\gamma_m = 0$. When $\lambda \to +\infty$, $\nu \to 0$ and holding safety stock and reserving capacity has no cost. $m_1 = \mu_1 = +\infty$. However $c < w$, therefore the optimal policy is to always deploy safety stock. Thus $\mu^* = 0$ and $\gamma_m = 1$.

Proof of Proposition 24
Proof. From the formulation,

\[ J_t(\mathbf{n}^t) = \max_{m_i^t \geq n_i^t} E_{\mathbf{d}, \mathbf{a}} \left\{ \max_{\Omega^t} \sum_{i=1}^k \left( \prod_{i=1}^k \sum_{t=1}^{m_i^t} x_{ij}^t a_{ij}^t - \ell_i m_i^t \right) + \beta E_{s} \left[ J_{t+1}(n_{1}^{t+1}, \ldots, n_{k}^{t+1}) \right] \right\} \]

\[ + \sum_{i=1}^k \ell_i n_i^t \]

Define the corresponding total expected discounted profit function for the seller

\[ f(\mathbf{n}, \mathbf{a}^n, \mathbf{d}^n) = \max_{\Omega} \sum_{i=1}^k \sum_{j=1}^{n_i} \pi_i x_{ij} a_{ij} \]

\[ + \beta E_s \left[ J \left( \sum_{j=1}^{n_1} s_{1j}(\alpha_1(x_{1j})), \ldots, \sum_{j=1}^{n_k} s_{kj}(\alpha_k(x_{kj})) \right) \right] \]

\[ L(\mathbf{n}) = E_{\mathbf{a}^n, \mathbf{a}^t} \{ f(\mathbf{n}, \mathbf{a}^n, \mathbf{d}^n) \} \]

\[ J(\mathbf{n}) = \max_{m_i \geq n_i} \left\{ L(\mathbf{m}) - \sum_{i=1}^k \ell_i m_i \right\} + \sum_{i=1}^k \ell_i n_i \]

Here \( J(\mathbf{n}) \) is the seller’s equilibrium optimal profit to go function. Given \( \mathbf{n} \), we want to show that \( L(\mathbf{n}) \) is increasing in \( n_i \). Assume \( \mathbf{n}' = \mathbf{n} + \mathbf{e}_t \), where \( e_{tt} = 1 \) and \( e_{ij} = 0 \). For each sample path \( \mathbf{a}^n = (\mathbf{a}_1, \ldots, \mathbf{a}_t, \ldots, \mathbf{a}_k) \) and \( \mathbf{a}^n' = (\mathbf{a}_1, \ldots, \mathbf{a}_t', \ldots, \mathbf{a}_k) \), \( \mathbf{d}^n = (\mathbf{d}_1, \ldots, \mathbf{d}_t, \ldots, \mathbf{d}_k) \) and \( \mathbf{d}^n' = (\mathbf{d}_1, \ldots, \mathbf{d}_t', \ldots, \mathbf{d}_k) \) with \( \mathbf{a}_t' = (\mathbf{a}_t, a_{t,n_t+1}) \) and
\[ d'_t = (d_t, d_{t,n+1}). \]

\[
f(n, a^n, d^n) = \max \quad \sum_{i=1}^{k} \sum_{j=1}^{m} \pi_i x_{ij} a_{ij} \\
\sum_{i=1}^{k} \sum_{j=1}^{m} x_{ij} a_{ij} d_{ij} \leq K \\
+ \beta E_s \left[ J \left( \ldots, \sum_{j=1}^{n} s_{tj} (\alpha_t(x_{tj})), \ldots \right) \right] \\
\leq \max \quad \sum_{i=1}^{k} \sum_{j=1}^{m} \pi_i x_{ij} a_{ij} \\
\sum_{i=1}^{k} \sum_{j=1}^{m} x_{ij} a_{ij} d_{ij} \leq K \\
+ \beta E_s \left[ J \left( \ldots, \sum_{j=1}^{n} s_{tj} (\alpha_t(x_{tj})), s_{t,n+1} (\alpha_t(0)), \ldots \right) \right] \\
\leq \max \quad \sum_{i=1}^{k} \sum_{j=1}^{m'} \pi_i x_{ij} a_{ij} \\
\sum_{i=1}^{k} \sum_{j=1}^{m'} x_{ij} a_{ij} d_{ij} \leq K \\
x_{t,n+1} a_{t,n+1} = 0 \\
+ \beta E_s \left[ J \left( \ldots, \sum_{j=1}^{n'} s_{tj} (\alpha_t(x_{tj})), \ldots \right) \right] \\
\leq \max \quad \sum_{i=1}^{k} \sum_{j=1}^{m'} \pi_i x_{ij} a_{ij} \\
\sum_{i=1}^{k} \sum_{j=1}^{m'} x_{ij} a_{ij} d_{ij} \leq K \\
+ \beta E_s \left[ J \left( \ldots, \sum_{j=1}^{n'} s_{tj} (\alpha_t(x_{tj})), \ldots \right) \right] \\
= f(n', a^n, d^n)
\]

The first inequality is justified by the fact that \(\alpha(0) \geq 0\) and \(J(n)\) is increasing in \(n\); the second inequality is justified by that when \(a_{t,n+1} = 0, x_{t,n+1} = 1\); the
last inequality is justified by that maximization over large feasible set will get larger optimal solution. Therefore
\[ L(n) = E_{d,s}[f(n, a^n, d^n)] \]
\[ \leq E_{d,s}[f(n', a'^n, d'^n)] = L(n') \]

Thus \( L(n) \) is increasing in \( n_i \). Moreover,
\[ f(n, a, d) \leq \pi_1 K + \beta \pi_1 K + \beta^2 \pi_1 K + \ldots = \frac{1}{1 - \beta} \pi_1 K \]
\[ \implies 0 \leq L(n) \leq \frac{1}{1 - \beta} \pi_1 K \]

We have \( L(n) \) is increasing and converges to a finite constant when \( n \to \infty \). At the same time, \( -\sum_{i=1}^k \ell_i m_i \to -\infty \) when \( m_i \to \infty \). There exists \( \mathbf{m}_i \) so that for any \( \mathbf{m} \geq \mathbf{m} = (\mathbf{m}_1, \ldots, \mathbf{m}_k) \)
\[ L(\mathbf{m}) - \sum_{i=1}^k \ell_i m_i \leq L(\mathbf{0}) - \sum_{i=1}^k \ell_i 0 \]

Therefore, there exists an optimal solution \( \mathbf{n}^* = (n_1^*, \ldots, n_k^*) \) that maximizes \( L(\mathbf{m}) - \sum_{i=1}^k \ell_i m_i \). When the seller starts with no customer \( n_i = 0 \), the optimal policy for the dynamic infinite horizon model is always to keep a constant customer base size of \( n^* \) whenever the number of customers drops below it. \( \square \)

**Proof of Lemma 25**

*Proof.* When \( a_{ij} = 0 \), there is no order and \( q_{ij} = 0 \). For the capacity allocation optimization problem, we only consider \( a_{ij} = 1 \). Since \( \sum_{i=1}^k \sum_{j=1}^{n_i} \alpha_i(x_{ij}) \) is concave function and the feasible set \( \Omega \) is compact set, the optimal solution is guaranteed to exist. Moreover, there exists feasible \( \mathbf{x} \) so that \( \mathbf{x} > 0 \), based on the Kuhn-Tucker
Theorem under Convexity, we have $x^*$ is optimal if and only if it satisfies the following Kuhn-Tucker first-order condition:

$$\pi_i + \frac{\beta \ell_i \alpha'(x^*)}{d_{ij}} = \lambda - \mu_{ij} + \nu_{ij}$$

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} x^*_{ij} d_{ij} = K$$

$$\mu_{ij} x^*_{ij} = 0$$

$$\nu_{ij} (1 - x^*_{ij}) = 0$$

$$\lambda \geq 0, \mu_{ij} \geq 0, \nu_{ij} \geq 0$$

Proof of Proposition 26

Proof. From the proof of Lemma 1, we know that there exists $\lambda$ so that (1) $x^*_{ij} = 0$ if and only if $\alpha'_i(0) \leq d_{ij} \frac{\lambda - \pi_i}{\beta \ell_i}$ (2) $x^*_{ij} = 1$ if and only if $\alpha'_i(1) \geq d_{ij} \frac{\lambda - \pi_i}{\beta \ell_i}$ (3) $0 < x^*_{ij} = (\alpha'_i)^{-1} \left(d_{ij} \frac{\lambda - \pi_i}{\beta \ell_i}\right) < 1$ if and only if $\alpha'_i(1) < d_{ij} \frac{\lambda - \pi_i}{\beta \ell_i} < \alpha'_i(0)$. Apparently, the optimal fill rate $x^*_{ij}$ is independent of disruption risk $p_i$ and future demand volume $D_i$ once capacity occurs and customers availability $a_{ij}$ and order $d_{ij}$ is realized. In addition, $x^*_{ij}$ is decreasing in $d_{ij}$.

1. When $d_{ij}$ increases, if $\lambda$ decreases, $x^*_{ij}$ for all other $i$ and $j$ is increasing. As a result, $d_{ij}$ for all other $i$ and $j$ are increasing. If $x^*_{ij}$ is also increasing, the total capacity allocated $\sum_{i=1}^{k} \sum_{j=1}^{n_i} x^*_{ij} d_{ij}$ is increasing. This conflicts with the constraint $\sum_{i=1}^{k} \sum_{j=1}^{n_i} x^*_{ij} d_{ij} = K$. Therefore $x^*_{ij}$ must be decreasing. If $\lambda$ increases, it is easy to verify that $x^*_{ij}$ is also decreasing. Thus $x^*_{ij}$ is decreasing in $d_{ij}$.

2. When $\pi_i$ increases, if $\lambda$ decreases, $x^*_{ij}$ is increasing. If $\lambda$ increases, $x^*_{ij}$ for all other $i$ and $j$ are decreasing. As a result, the total capacity allocated to all
other customers is decreasing. Given the same total capacity $K$ and constant $d_{ij}$, $x_{ij}^*$ must be increasing. Thus $x_{ij}^*$ is increasing in $\pi_i$.

3. When $\ell_i$ increases, if $\lambda$ decreases, $x_{ij}^*$ is increasing. If $\lambda$ increases, $x_{ij}^*$ for all other $i$ and $j$ are decreasing. As a result, the total capacity allocated to all other customers is decreasing. Given the same total capacity $K$ and constant $d_{ij}$, $x_{ij}^*$ must be increasing. Thus $x_{ij}^*$ is increasing in $\ell_i$.

4. When $\alpha'_i()$ is increasing, $(\alpha'_i)^{-1}$ is decreasing. If $\lambda$ decreases, $x_{ij}^*$ is increasing. If $\lambda$ increases, $x_{ij}^*$ for all other $i$ and $j$ are decreasing. As a result, the total capacity allocated to all other customers is decreasing. Given the same total capacity $K$ and constant $d_{ij}$, $x_{ij}^*$ must be increasing. Thus $x_{ij}^*$ is increasing in $\alpha'_i()$.

Proof of Proposition 27

Proof. Under logistic retention function $\alpha_i(x) = \frac{1}{1+exp(-ax)}$, 
\[
\frac{\partial \alpha_i(x)}{\partial x} = \frac{a * exp(-ax)}{(1+exp(-ax))^2}
\]
\[
\frac{\partial^2 \alpha_i(x)}{\partial a \partial x} = \frac{((1-a^2) + (1+a^2) exp(-ax)) exp(-ax)}{(1+exp(-ax))^3}
\]

Take the derivative of the above, we have 
\[
-2a + 2a * exp(-ax) - a (1 + a^2) exp(-ax) = a ((1-a^2) exp(-ax) - 2) < 0
\]

Thus $\alpha'_i(x)$ is concave with respect to $a$. When $a = 0$, $\frac{\partial \alpha'_i(x)}{\partial a} > 0$ and thus $\alpha'_i(x)$ is first increasing. When $a \rightarrow +\infty$, $\frac{\partial \alpha'_i(x)}{\partial a} < 0$ and thus $\alpha'_i(x)$ is decreasing in $a$. Thus the fill rate $x_{ij}^*$ is first increasing then decreasing in $a$. 

138
Under exponential retention function \( \alpha_i(x) = 1 - \exp(-ax) \), we have

\[
\frac{\partial \alpha_i(x)}{\partial x} = a \cdot \exp(-ax)
\]

\[
\frac{\partial a \cdot \exp(-ax)}{\partial a} = \exp(-ax) - a^2 \exp(-ax)
\]

Similarly, \( \alpha_i'(x) \) is increasing when \( a \) is small and decreasing when \( a \) is big. Thus the fill rate \( x_{ij}^* \) is also first increasing then decreasing in \( a \). \( \square \)

**Proof of Lemma 28**

**Proof.** We start the proof for \( k = 2 \). Define

\[
g(m_1, m_2) (d_{1,1}, \ldots, d_{1,m_1}, d_{2,1}, \ldots, d_{2,m_2}) = \max \Omega \sum_{i=1}^{k} \left\{ \sum_{j=1}^{m_i} \pi_i x_{ij} d_{ij} - \beta \ell_i \alpha_i(x_{ij}) \right\}
\]

Then we have that

\[
G(m_1, m_2) = E_a \left[ g(m_1, m_2) (d_{1,1}, \ldots, d_{1,m_1}, d_{2,1}, \ldots, d_{2,m_2}) \right]
\]

First, we prove that \( g(m_1 + 1, m_2) - g(m_1, m_2) \) is non-increasing in \( n \).

\[
g(m_1 + 1, m_2) (d_{1,1}, \ldots, d_{1,m_1+1}, d_{2,1}, \ldots, d_{2,m_2})
\]

\[
- g(m_1, m_2) (d_{1,1}, \ldots, d_{1,m_1}, d_{2,1}, \ldots, d_{2,m_2})
\]

\[
= g(m_1 + 1, m_2) (d_{1,1}, \ldots, d_{1,m_1+1}, d_{2,1}, \ldots, d_{2,m_2})
\]

\[
- g(m_1 + 1, m_2) (d_{1,1}, \ldots, d_{1,m_1}, \infty, d_{2,1}, \ldots, d_{2,m_2})
\]

\[
+ g(m_1 + 1, m_2) (d_{1,1}, \ldots, d_{1,m_1}, \infty, d_{2,1}, \ldots, d_{2,m_2})
\]

\[
- g(m_1, m_2) (d_{1,1}, \ldots, d_{1,m_1}, d_{2,1}, \ldots, d_{2,m_2})
\]

Easy to check

\[
g(m_1 + 1, m_2) (d_{1,1}, d_{1,2}, \ldots, d_{1,m_1}, \infty, d_{2,1}, d_{2,2}, \ldots, d_{2,m_2})
\]

\[
- g(m_1, m_2) (d_{1,1}, d_{1,2}, \ldots, d_{1,m_1}, d_{2,1}, d_{2,2}, \ldots, d_{2,m_2})
\]

\[
= \pi_1 \left( K - \sum_{j=1}^{m_1} d_{1j} - \sum_{j=1}^{m_2} d_{2j} \right) + \beta \ell_1 \alpha_1 (0)
\]
\[ g_{(m_1+1,m_2)} (d_{1,1}, d_{1,2}, ..., d_{1,m_1+1}, d_{2,1}, d_{2,2}, ..., d_{2,m_2}) \]

\[ -g_{(m_1+1,m_2)} (d_{1,1}, d_{1,2}, ..., d_{1,m_1}, \infty, d_{2,1}, d_{2,2}, ..., d_{2,m_2}) \]

\[ = - \int_{d_{1,m_1+1}}^{\infty} \beta \ell_1 \alpha_1 \left( \frac{q^* (d_{1,1}, d_{1,2}, ..., w, d_{2,1}, d_{2,2}, ..., d_{2,m_2})}{w} \right) \]

\[ \frac{q^* (d_{1,1}, d_{1,2}, ..., w, d_{2,1}, d_{2,2}, ..., d_{2,m_2})}{w^2} \]

\[ = \beta \ell_1 \int_{d_{1,1}}^{\infty} \alpha_1 \left( \frac{q^* (w, d_{1,2}, ..., d_{1,m_1+1}, d_{2,1}, d_{2,2}, ..., d_{2,m_2})}{w} \right) \]

\[ \frac{q^* (w, d_{1,2}, ..., d_{1,m_1+1}, d_{2,1}, d_{2,2}, ..., d_{2,m_2})}{w^2} \]

So

\[ G (m_1 + 1, m_2) - G (m_1, m_2) \]

\[ = E_d \left[ g_{(m_1+1,m_2)} (d_{1,1}, ..., d_{1,m_1+1}, d_{2,1}, ..., d_{2,m_2}) \right] \]

\[ - E_d \left[ g_{(m_1,m_2)} (d_{1,1}, ..., d_{1,m_1}, d_{2,1}, ..., d_{2,m_2}) \right] \]

Since \( \pi_1 \left( K - \sum_{j=1}^{m_1} d_{1j} - \sum_{j=1}^{m_2} d_{2j} \right)^+ \) is non-increasing in \( m_2 \),

\[ E_d \left[ \pi_1 \left( K - \sum_{j=1}^{m_1} d_{1j} - \sum_{j=1}^{m_2} d_{2j} \right)^+ \right] \] is decreasing in \( m_2 \). For the same order quantity from the customer \( w \),

\( q^* (w, ..., d_{1,m_1+1}, d_{2,1}, ..., d_{2,m_2}) \) is non-increasing in \( m_2 \), thus

\( q^* (w, ..., d_{1,m_1+1}, d_{2,1}, ..., d_{2,m_2}) \) is non-increasing in \( m_2 \). Based on our assumption that \( \alpha'_1 (y) \) is increasing in \( y \), we have

\[ \alpha'_1 \left( \frac{q^* (w, ..., d_{1,m_1+1}, d_{2,1}, ..., d_{2,m_2})}{w} \right) q^* \left( \frac{w, ..., d_{1,m_1+1}, d_{2,1}, ..., d_{2,m_2}}{w^2} \right) \]

is non-increasing in \( m_2 \). Finally, we have \( G (m_1 + 1, m_2) - G (m_1, m_2) \) is non-
increasing in $m_2$

\[
J(m_1 + 1, m_2) - J(m_1, m_2)
\]

\[
= \frac{1}{1 - \beta} E_a \left[ G \left( \sum_{j=1}^{m_1 + 1} a_{1j}, \sum_{j=1}^{m_2} a_{2j} \right) - G \left( \sum_{j=1}^{m_1} a_{1j}, \sum_{j=1}^{m_2} a_{2j} \right) \right]
\]

\[- \frac{\beta}{1 - \beta} \ell_1 \alpha_1 (1) p_1 - \ell_1
\]

Next, To prove $J(m_1, m_2)$ is sub-modular, we need to prove that $J(m_1 + 1, m_2) - J(m_1, m_2)$ is non-increasing in $m_2$. Given $G(m_1 + 1, m_2) - G(m_1, m_2)$ is non-increasing in $m_2$, we need to show

\[
E_a \left[ G \left( \sum_{j=1}^{n_1 + 1} a_{1j}, \sum_{j=1}^{n_2} a_{2j} \right) - G \left( \sum_{j=1}^{n_1} a_{1j}, \sum_{j=1}^{n_2} a_{2j} \right) \right]
\]
is non-increasing in $n_2$

$$E_a \left[ G \left( \sum_{j=1}^{n_1} a_{1j}, \sum_{j=1}^{n_2} a_{2j} \right) - G \left( \sum_{j=1}^{n_1} a_{1j}, \sum_{j=1}^{n_2} a_{2j} \right) \right]$$

$$= (1 - p_1) E_a \left[ G \left( \sum_{j=1}^{n_1} a_{1j} + 1, \sum_{j=1}^{n_2} a_{2j} \right) - G \left( \sum_{j=1}^{n_1} a_{1j}, \sum_{j=1}^{n_2} a_{2j} \right) \right]$$

$$E_a \left[ G \left( \sum_{j=1}^{n_1+1} a_{1j}, \sum_{j=1}^{n_2} a_{2j} \right) - G \left( \sum_{j=1}^{n_1} a_{1j}, \sum_{j=1}^{n_2} a_{2j} \right) \right]$$

$$- E_a \left[ G \left( \sum_{j=1}^{n_1+1} a_{1j}, \sum_{j=1}^{n_2} a_{2j} \right) - G \left( \sum_{j=1}^{n_1} a_{1j}, \sum_{j=1}^{n_2} a_{2j} \right) \right]$$

$$= (1 - p_1) E_a \left[ G \left( \sum_{j=1}^{n_1} a_{1j} + 1, \sum_{j=1}^{n_2} a_{2j} \right) - G \left( \sum_{j=1}^{n_1} a_{1j}, \sum_{j=1}^{n_2} a_{2j} \right) \right]$$

$$- (1 - p_1) E_a \left[ G \left( \sum_{j=1}^{n_1} a_{1j} + 1, \sum_{j=1}^{n_2} a_{2j} \right) - G \left( \sum_{j=1}^{n_1} a_{1j}, \sum_{j=1}^{n_2} a_{2j} \right) \right]$$

$$= (1 - p_1) (1 - p_2) E_a \left[ G \left( \sum_{j=1}^{n_1} a_{1j} + 1, \sum_{j=1}^{n_2} a_{2j} + 1 \right) + G \left( \sum_{j=1}^{n_1} a_{1j}, \sum_{j=1}^{n_2} a_{2j} \right) \right]$$

$$- (1 - p_1) (1 - p_2) E_a \left[ G \left( \sum_{j=1}^{n_1} a_{1j}, \sum_{j=1}^{n_2} a_{2j} + 1 \right) + G \left( \sum_{j=1}^{n_1} a_{1j} + 1, \sum_{j=1}^{n_2} a_{2j} \right) \right]$$

$$\leq 0$$

If there are $k > 2$ types of customers available define

$$g(n_1, n_2, ..., n_k) (d) = \max_{\sum_{i=1}^k \sum_{j=1}^{n_j} q_{ij} \leq K} \sum_{i=1}^k \sum_{j=1}^{n_j} \pi_i q_{ij} - \beta \sum_{i=1}^k \ell_i \alpha_i \left( \frac{q_{ij}}{d_{ij}} \right)$$

Then $J(n)$ can be proved to be sub-modular under the same constraints, using the same method.

**Proof of Proposition 29**

142
Proof. Define

\[ g(d_1, d_2, ..., d_t) = \min_{0 \leq x_j \leq 1} \left( \sum_{j=1}^{t} \alpha(x_j) \right) \]
\[ \sum_{j=1}^{t} x_j d_j \leq K \]

\[ g(t) = E_d \left[ g(d_1, d_2, ..., d_t) \right] \]

\[ G(n) = \sum_{t=0}^{n} C_n^t p^{n-t}(1-p)^t g(t) = E_a \left[ g \left( \sum_{j=1}^{n} a_j \right) \right] \]

So we have

\[ J(n) = \frac{R(n)}{1 - \beta} - \ell n \frac{1 + \beta \left( p \bar{\alpha} (1) - 1 \right)}{1 - \beta} - \frac{\ell \beta}{1 - \beta} G(n) \]

Here \( R(n) = \pi E \left[ \sum_{j=1}^{n} a_j d_j \wedge K \right] \). When there is only one type of customer, \( J(n) \) is concave and \( n^* \) satisfies

\[ n^* = \min \left\{ n \geq 0 \mid J(n+1) - J(n) \leq 0 \right\} \]

Obvious to conclude that \( n^* \) is increasing in \( J(n+1) - J(n) \)

\[ J(n+1) - J(n) = \frac{R(n+1) - R(n)}{1 - \beta} - \ell \frac{1 + \beta \left( p \bar{\alpha} (1) - 1 \right)}{1 - \beta} \]
\[ - \frac{\ell \beta (1-p)}{1 - \beta} \left\{ E_a \left[ g \left( \sum_{i=1}^{n} a_i + 1 \right) - g \left( \sum_{i=1}^{n} a_i \right) \right] \right\} \]

When \( \pi \) increases, we have \( \frac{R(n+1) - R(n)}{1 - \beta} \) is increasing in \( \pi \) and \( J(n+1) - J(n) \) is increasing in \( \pi \), therefore \( n^* \) is increasing in \( \pi \). Since \( \frac{1 + \beta \left( p \bar{\alpha} (1) - 1 \right)}{1 - \beta} > 1 - \beta = 1 \), the second term and third term of \( J(n+1) - J(n) \) are decreasing \( \ell \), thus \( n^* \) is decreasing in \( \ell \). The proof for parameter \( K \) is more complicated. Since function \( a \wedge y \) is super-modular in \( (a, y) \), therefore \( R(n+1) - R(n) \) is increasing in \( K \). Moreover, we have

\[ g(n+1) - g(n) = -E_d \left[ \int_{\mathbb{R}}^{\infty} \alpha' \left( \frac{q_1^*(w, d_2, ..., d_n, d_{n+1})}{w} \right) q_1^*(w, d_2, ..., d_n, d_{n+1}) \frac{dw}{w^2} \right] + \bar{\alpha}(0) \]
For higher $K$, the fill rate $q^*_1(w, d_2, \ldots, d_n, d_n) + 1$ is higher, therefore $g(n+1) - g(n)$ is decreasing in $K$, thus $E_a[g(\sum_{i=1}^n a_i + 1) - g(\sum_{i=1}^n a_i)]$ can be proved to be decreasing in $K$. Thus $J(n + 1) - J(n)$ is increasing in $K$ and $n^*$ is increasing in $K$. 

**Proof of Proposition 30**

**Proof.** To see the impact of $D$ over the optimal customer base size, we need to verify the impact of $D$ over $J(n + 1) - J(n)$.

$$J(n + 1) - J(n) = \frac{1-p}{1-\beta} E_a \left[ G \left( \sum_{j=1}^n a_j + 1 \right) - G \left( \sum_{j=1}^n a_j \right) \right] - \frac{\beta}{1-\beta} \ell(1)p - \ell$$

Here $G(n)$ is defined as in the section of Optimal Acquisition Strategy. To show the monotonicity of $J(n + 1) - J(n)$ is equivalent to show that of $G(n + 1) - G(n)$.

For each sample path of $(d_1, \ldots, d_{n+1})$, we have

$$g_{n+1}(d_1, \ldots, d_{n+1}) - g_n(d_1, \ldots, d_n) = g_{n+1}(d_1, \ldots, d_{n+1}) - g_{n+1}(d_1, \ldots, d_n, \infty) + g_{n+1}(d_1, \ldots, d_n, \infty) - g_n(d_1, \ldots, d_n)$$

$$g_{n+1}(d_1, \ldots, d_n, \infty) - g_n(d_1, \ldots, d_n) = \pi \left( K - \sum_{j=1}^n d_j \right)^+ - \beta \ell \alpha(0)$$

$$g_{n+1}(d_1, \ldots, d_{n+1}) - g_{n+1}(d_1, \ldots, d_n, \infty) = - \int_{w=d_{n+1}}^\infty \frac{\hat{g}(d_1, d_2, \ldots, d_n, w)}{\hat{\omega}w} dw$$

$$= - \int_{w=d_{n+1}}^\infty (\pi - \lambda) \frac{q^*_1(d_1, d_2, \ldots, d_n, w)}{w} dw$$
\[ G(n+1) - G(n) \]
\[ = E_d \left[ \int_{w=d_{n+1}}^{\infty} \alpha' \left( \frac{q_{n+1}^* (d_1, d_2, \ldots, d_n, w)}{w} \right) q_{n+1}^* (d_1, d_2, \ldots, d_n, w) dw \right] \]
\[ + E_d \left[ \pi \left( K - \sum_{j=1}^{n} d_j \right)^+ \right] - \beta \ell \bar{\pi} (0) \]

It is easy to verify that when \( d_j \sim D \) increases in convex order,
\[ E_d \left[ \pi \left( K - \sum_{j=1}^{n} d_j \right)^+ \right] \] is increasing. Similarly, we can show that the first term is increasing when \( d_j \sim D \) increases in convex order. Thus \( G(n+1) - G(n) \) is increasing when \( D \) increases in convex order. As a result, \( J(n+1) - J(n) \) is increasing when \( D \) increases in convex order. Therefore \( n^* \) is increasing when \( D \) increases in convex order. \( \square \)

**Proof of Proposition 31**

**Proof.** According the following formula for \( J(n) \), we have
\[ \frac{\partial J(n+1) - J(n)}{\partial p} = \frac{1}{1-\beta} \frac{\partial R(n+1) - R(n)}{\partial p} - \ell \frac{\beta \bar{\pi} (1)}{1-\beta} \]
\[ + \frac{\ell \beta}{1-\beta} \frac{\partial (p-1) E_a \left[ g \left( \sum_{j=1}^{n} a_j + 1 \right) - g \left( \sum_{j=1}^{n} a_j \right) \right]}{\partial p} \]

Since
\[ \frac{\partial (p-1) E_a \left[ g \left( \sum_{j=1}^{n} a_j + 1 \right) - g \left( \sum_{j=1}^{n} a_j \right) \right]}{\partial p} \]
\[ = E_a \left[ g \left( \sum_{j=1}^{n} a_j + 1 \right) - g \left( \sum_{j=1}^{n} a_j \right) \right] - (1-p) \frac{\partial E_a \left[ g \left( \sum_{j=1}^{n} a_j + 1 \right) - g \left( \sum_{j=1}^{n} a_j \right) \right]}{\partial p} \]

Since \( g(n+1) - g(n) \) is decreasing in \( n \) and \( \sum_{j=1}^{n} a_j \) is stochastically decreasing
when \( p \) increases. According to the property of stochastic order, we have

\[
\frac{\partial E_a \left[ g \left( \sum_{j=1}^{n} a_j + 1 \right) - g \left( \sum_{j=1}^{n} a_j \right) \right]}{\partial p} < 0
\]

Thus \( \frac{\partial^{(p-1)}E_a [g(\sum_{j=1}^{n} a_j+1) - g(\sum_{j=1}^{n} a_j)]}{\partial p} > 0 \). Similarly, we can show that \( \frac{\partial R(n+1)-R(n)}{\partial p} < 0 \). Further, take the second order derivative of \( J(n+1) - J(n) \) by writing out the dependency of \( R(n+1) - R(n) \) and \( E_a \left[ g \left( \sum_{j=1}^{n} a_j + 1 \right) - g \left( \sum_{j=1}^{n} a_j \right) \right] \) over \( p \), we have \( \frac{\partial J(n+1)-J(n)}{\partial p^2} < 0 \). Therefore, there exists \( p^* \) so that \( J(n+1) - J(n) \) is increasing in \( p \) when \( p < p^* \), decreasing in \( p \) when \( p \geq p^* \). Therefore, \( n^* \) is increasing in \( p \) when \( p < p^* \), decreasing in \( p \) when \( p \geq p^* \). \( \square \)

*Proof of Proposition 32*
Proof. Under the retention function $\alpha(x, a)$, we have

$$J(n + 1) - J(n) = \frac{R(n + 1) - R(n)}{1 - \beta} - \frac{r_1 + \beta (p\alpha (1) - 1)}{1 - \beta}$$

$$- \frac{\ell \beta}{1 - \beta} E_{d,a} \left[ g \left( \sum_{i=1}^{n+1} a_i \right) - g \left( \sum_{i=1}^{n} a_i \right) \right]$$

$$\frac{\partial J(n + 1) - J(n)}{\partial a} = \frac{\ell \beta}{1 - \beta} \frac{\partial \alpha (1, a)}{\partial a} - \frac{\ell \beta}{1 - \beta} (1 - p)$$

$$\frac{\partial E_{d,a} \left[ g \left( \sum_{i=1}^{n} a_i + 1 \right) - g \left( \sum_{i=1}^{n+1} a_i \right) \right]}{\partial a}$$

$$= \frac{\ell \beta}{1 - \beta} \frac{\partial \alpha (1, a)}{\partial a} - \frac{\ell \beta}{1 - \beta} (1 - p)$$

$$\sum_{i=0}^{n} C_{n}^{j} p^{n-j} (1 - p)^j \frac{\partial}{\partial a} \left( g (j + 1) - g (j) \right)$$

$$\frac{\partial}{\partial a} \left( g (j + 1) - g (j) \right) = \frac{\partial}{\partial a} \left[ \int_{W=d_1}^{\infty} \alpha' \left( \frac{q^*_{1} (w, d_2, d_3, d_{j+1})}{w}, a \right) q^*_{1} (w, d_2, d_3, d_{j+1}) dw \right]$$

$$- \frac{\partial \alpha (0, a)}{\partial a}$$

$$= - \int_{W=d_1}^{\infty} \frac{\partial \alpha' \left( \frac{q^*_{1} (w, d_2, d_3, d_{j+1})}{w}, a \right)}{\partial a} q^*_{1} (w, d_2, d_3, d_{j+1}) dw$$

$$- \frac{\partial \alpha (0, a)}{\partial a} < 0$$

$$\frac{\partial J(n + 1) - J(n)}{\partial a} > 0$$

Therefore, $n^*$ is increasing in $a$ when $\frac{\partial \alpha'(x, a)}{\partial a} > 0$.

Proof of Proposition 33

Proof. In period 2, the supplier has no incentive to retain customers. Therefore, the optimal capacity allocation scheme does not differentiate the customers at all. It is
optimal for the supplier to satisfy the customers’ order as much as possible. Then
\[
\max_{\Omega} \left( \pi \sum_{j=1}^{m^2} x_j^2 d_j^2 a_j^2 - \ell \left( m^2 - n^2 \right) \right) = \pi \left( K - \left( K - \sum_{j=1}^{m^2} a_j^2 d_j^2 \right)^+ \right) - \ell \left( m^2 - n^2 \right)
\]

Therefore,
\[
J_2 (n^2) = \pi K + \ell n^2 - \min_{m^2 \geq n^2} R^2 (m^2)
\]
\[
R^2 (m^2) = E_{d,a} \left[ \pi \left( K - \sum_{j=1}^{m^2} a_j^2 d_j^2 \right)^+ \right] + \ell n^2
\]

It is easy to show that \( R^2 (m^2) \) is convex with respect to \( m^2 \). Therefore, there exists \( n^{2*} \) so that the optimal customer base policy in period 2 is an add-up-to policy with add-up-to level \( n^{2*} \). Furthermore,
\[
J_2 (n^2) = \begin{cases} 
\pi K + \ell n^2 - R^2 (n^{2*}) \\
\pi K - E_{d,a} \left[ \pi \left( K - \sum_{j=1}^{n^2} a_j^2 d_j^2 \right)^+ \right] 
\end{cases} , \ n^2 \leq n^{2*}
\]
\[
J_2 (n^2) = \begin{cases} 
\pi K - E_{d,a} \left[ \pi \left( K - \sum_{j=1}^{n^2} a_j^2 d_j^2 \right)^+ \right] \\
\pi K + \ell n^2 - R^2 (n^{2*}) 
\end{cases} , \ n^2 > n^{2*}
\]

Thus \( J_2 (n^2) \) is increasing and concave in \( n^2 \). Further, consider the capacity allocation scheme in period 1, it is equivalent to solve
\[
\max_{\Omega^1} \left( \pi \sum_{j=1}^{m^1} x_j^1 d_j^1 a_j^1 \right) + \beta E_s \left[ J_2 (n^2) \right] - \ell (m^1 - n^1)
\]
\[
\iff \pi \left( K - \left( K - \sum_{j=1}^{m^1} a_j^1 d_j^1 \right)^+ \right) + \beta \max_{\Omega^1} E_s \left[ J_2 (n^2) \right] - \ell (m^1 - n^1)
\]

At last, consider the optimal customer base policy in period 1,
\[
J_1 (n^1) = \pi K + \ell n^1 - \min_{m^1 \geq n^1} R^1 (m^1)
\]
\[
R^1 (m^1) = R^2 (m^1) - \beta \max_{\Omega^1} E_s \left[ J_2 \left( \sum_{j=1}^{m^1} s_j^1 (x_j^1) \right) \right]
\]

148
Still, we have \( R^2 (m^1) \) is convex in \( m^1 \) and \( \beta \max_{\Omega^1} E_s \left[ J_2 \left( \sum_{j=1}^{m^1} s^1_j \left( \alpha \left( x^1_j \right) \right) \right) \right] \) is increasing. There exists unique optimal \( m^{1*} \) that maximizes \( R^1 (m^1) \) and the optimal customer base policy in period 1 is still an add-up-to policy with add-up-to level \( m^{1*} \). Consider \( m^{1*} (\beta) \) and \( m^{2*} \), the optimal \( m^{1*} (\beta) \) should satisfy that it is the first one that

\[
\Delta R^2 (m^1) - \beta \max_{\Omega^1} E_s \left[ J_2 \left( \sum_{j=1}^{m^1} s^1_j \left( \alpha \left( x^1_j \right) \right) \right) \right] \geq 0.
\]

As we can see \( \Delta \max_{\Omega^1} E_s \left[ J_2 \left( \sum_{j=1}^{m^1} s^1_j \left( \alpha \left( x^1_j \right) \right) \right) \right] \) is positive, therefore, the above equation is decreasing in \( \beta \). Thus \( m^{1*} (\beta) \) is increasing in \( \beta \) with \( m^{1*} (0) = m^{2*} \)

\[ \square \]

**Proof of Proposition 34**

Proof. In period 2, the supplier has no incentive to retain customers. Therefore, the optimal capacity allocation scheme does not differentiate the customers at all. It is optimal for the supplier to satisfy the customers’ order as much as possible. For the optimal allocation in period 1, since \( J_2 () \) is increasing and concave function, thus the optimal capacity allocation policy should satisfy

\[
\max_{\Omega^1} E_s \left[ J_2 \left( \sum_{j=1}^{m^1} s^1_j \left( \alpha \left( x^1_j \right) \right) \right) \right] \Leftrightarrow \max_{\Omega^1} \sum_{j=1}^{m^1} s^1_j \left( \alpha \left( x^1_j \right) \right)
\]

in stochastic order

However, from the results of Boland (2006), it is not equivalent to solve

\[
\max_{\Omega^1} E \left[ \sum_{j=1}^{m^1} s^1_j \left( \alpha \left( x^1_j \right) \right) \right] \Leftrightarrow \max_{\Omega^1} \sum_{j=1}^{m^1} \alpha \left( x^1_j \right)
\]

When \( \beta > 0 \), the optimal capacity allocation policy does not depend on the probability of whether the company makes to the second round. It is always optimal to
retain as many customers as possible.

Proof of Proposition 35

Proof. Similarly, at period 2, the supplier has no incentive to sacrifice profit for retaining customers. So its optimal capacity allocation policy is to satisfy the orders from customers of type 1 first given \( \pi_1 \geq \pi_2 \). If there is capacity left-over, then satisfy the orders from customers of type 2. Thus

\[
\max_{\Omega^2} \sum_{i=1}^{2} \left( \pi_i \sum_{j=1}^{m_i^2} x_{ij}^2 a_{ij}^2 - \ell_i \left( m_i^2 - n_i^2 \right) \right) \\
= \pi_1 K - (\pi_1 - \pi_2) \left( K - \sum_{j=1}^{m_1^2} a_{1j}^2 d_{1j}^2 \right)^+ \\
- \pi_2 \left( K - \sum_{j=1}^{m_2^2} a_{2j}^2 d_{2j}^2 \right)^+ - \sum_{j=1}^{m_1^2} \ell_i (m_i^2 - n_i^2)
\]

Similarly, we can formulate the optimal profit function as

\[
J_2 (n^2) = \pi_1 K + \sum_{i=1}^{2} \ell_i n_i^2 - \min_{m_i^2 \geq n_i^2} R (m)
\]

\[
R (m) = E_{d,a} \left[ (\pi_1 - \pi_2) \left( K - \sum_{j=1}^{m_1^2} a_{1j}^2 d_{1j}^2 \right)^+ + \pi_2 \left( K - \sum_{j=1}^{m_2^2} a_{2j}^2 d_{2j}^2 \right)^+ \right] \\
+ \sum_{i=1}^{2} \ell_i m_i^2
\]

From the formulation of \( R (m) \), we know that \( R (m) \) is sub-modular and jointly concave with respect to \( m \). In addition, we have \( J_2 (n^2) \) is increasing in \( n^2 \). Further, consider the capacity allocation scheme in period 1, we have

\[
\max_{\Omega^1} \sum_{i=1}^{2} \left( \pi_i \sum_{j=1}^{m_i^1} x_{ij}^1 a_{ij}^1 \right) + \beta E_s \left[ J_2 (n_1^2, n_2^2) \right]
\]
From the formulation, we know that when the customer demand $D_i^2$ increases in stochastic order, $J_2 (n_1^2, n_2^2)$ is increasing in $n_i^2$. Therefore for the same order quantity realization $d_{ij}^1$, the fill rate is increasing in customer’s orders $D_i^2$. Similarly, we can show the results for $p_i$ and other parameters. The details are omitted here. □
Bibliography


Robinson, L. W. (1990), Appropriate inventory policies when service affects future demands, vol. 88, Cornell University, Johnson Graduate School of Management.


Biography

Lu Huang was born in June 2, 1990 in Ji’an, jiangxi, China. Ms. Huang received her B.S. degree in Applied Mathematics at Nanjing University, Nanjing, China in 2010. She joined the Fuqua School of Business, Duke University as a Ph.D student in Operations Management in 2010.

From 2011-2012, she worked as vice president at Duke Chinese Students and Scholars Association. She also interned as a research scientist at Amazon Website Services in summer 2014, when she worked on inventory planning strategy under price fluctuation for assemble-to-order system.

She studies supplier risk management problem with Professor Jeannette Song. Her research interest includes:

- Global supply chain risk management
- Disaster relief and humanitarian supply chain management
- Mechanism design for supply chain sustainability