# DUAL-CHANNEL SOURCING AND SELLING STRATEGIES IN OPERATIONS MANAGEMENT

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF MANAGEMENT
SCIENCE AND ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Chen Peng
August 2011

This dissertation is online at: http://purl.stanford.edu/kz025bk4803

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

**Feryal Erhun Oguz, Primary Adviser**

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

**Hau Lee, Primary Adviser**

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

**Warren Hausman**

Approved for the Stanford University Committee on Graduate Studies.

**Patricia J. Gumport, Vice Provost Graduate Education**

*This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.*

# Abstract

The ability of manufacturers and suppliers to adapt to changing market conditions is crucial in today's uncertain business environment. Having more than one sourcing or selling channel with complementary services can be an effective strategy for firms to enhance their operational flexibility. This dissertation thus investigates how firms can utilize multiple channels to efficiently procure their production and service capacity or distribute sales volumes to meet the needs of a dynamic market. It contains two major parts:

First, in Chapters 2 and 3, we focus on the sourcing side and study how firms in capital-intensive industries can reduce their idle capacity while maintaining a high service level by purchasing production capacity from two supply sources. We construct a dual-mode equipment procurement model (DMEP), in which an equipment supplier provides two delivery modes to a firm: a base mode that is less expensive but slower and a flexible mode that is faster but more expensive. The combination of these two modes provides the firm the flexibility to mitigate demand risk at a potentially lower cost. Chapter 2 presents our theoretical approach and investigates a dynamic dual-source capacity expansion problem with consecutive leadtimes and demand backlogging. We demonstrate that the flexible orders follow a state-dependent base-stock policy; the base orders, however, follow only a partial-base-stock policy, which lacks structure and is difficult to track. Chapter 3 then tackles this problem from a practical perspective. Compromising optimality for applicability and efficiency, we construct a general DMEP heuristic that consists of three layers: a contract negotiation layer, in which the firm chooses the best combination of leadtime and price for each supply mode from the supply contract menu; a reservation layer, in which the firm reserves total equipment procurement quantities through the two

supply modes by paying the supplier a reservation fee up front before the planning horizon starts; and an execution layer, in which the firm acquires the latest demand information in each period and orders equipment through both supply modes. We numerically quantify the value of the added flexibility for the firm and explore how the optimal reservation and execution decisions would change with respect to the key model parameters.

Second, in Chapter 4, we instead study the selling side and discuss how a large commodity supplier should strategically allocate his limited production capacity between a fixed-price contract channel and a spot market to maximize his total sales income. We discuss two settings: one in which the equilibrium spot price follows an exogenous random distribution and one in which the equilibrium spot price is endogenously determined by the spot demand curve and the spot supply curve, both of which can be affected by the supplier's capacity allocation decision. In the former case, we find that the demand-price correlation and a risk-averse attitude are two reasons for the supplier to adopt a dual-channel strategy. The supplier should allocate more quantity to the spot channel if the contract channel demand and the spot price are more positively correlated, and he should allocate more to the contract channel if he is more risk-averse. In the latter setting, which further contains a contract trading stage and a spot trading stage, we show that a dual-channel policy is optimal in the first stage if the shifting effect of the supplier's spot allocation quantity on the default supply curve is stronger than the shifting effect of the unfulfilled contract channel demand on the default demand curve. Further, we demonstrate that it is not necessarily optimal to sell all leftover quantities in the spot market during the second stage. Using benchmark industry data, we quantify the average improvement in profit of adopting a dual-channel strategy versus using a single contract channel or a single spot channel through numerical analysis.

# Acknowledgements

First, I would like to recognize my principal advisor, Professor Feryal Erhun, for her outstanding instruction and support throughout my doctoral study at Stanford. Professor Erhun has been not only a great advisor and role model, but also a great mentor and friend of mine. She introduced me to my major dissertation topic; she guided me when I got confused with either research or life in general; she challenged me when I was content with the status quo and reluctant to probe further for a better outcome; she cheered for me when I demonstrated progress in my pursuit after an arduous devotion. I shall always remember the academic journey we took together.

I would also like to thank my co-advisor, Professor Hau Lee, for the guidance and help he provided me during these years. As a leader in our field, Professor Lee influenced me deeply not only through his incisive understanding of both the academia and the industry, but also through his generosity and kindness to others. Professor Lee also taught me to be a researcher who cares about both the theoretical rigor of his work and the positive impacts he can generate to the industry and the society.

I also want to sincerely thank Professor Warren Hausman for providing valuable feedback on my dissertation as a reading committee member and for being a mentor for me during my time at Stanford. His advice was instrumental for both my research and my professional career.

Finally, I would like to express special thanks to Zizhuo Wang, Yanchong Zheng, Tim Kraft, Sechan Oh, and Yichuan Ding, for all the help and advice they have offered me; to Karl Kempf, Erik Hertzler, and other members of Intel TMG, for their support and collaboration during the research project between Stanford and Intel; and to my dear classmates, Jessica McCoy, Hugo Mora, and Danny Greenia, for going through this challenging process together with me as my best friends.

# Dedication

I would like to dedicate this effort to my wife, Wenyi Cai, and my parents, Jianguo Peng and Gonglian Zhu. Their continuous support, encouragement, sacrifices, and love provided the foundation upon which this dissertation was achieved.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

## 1.1. Background

We are in a world where technology advancement, internet, and globalization make the entire business environment more complex and unpredictable than ever. Aligning supply with demand becomes a more difficult balancing act as uncertainties prevail in the global supply chain network. Consequently, firms in both high-tech and traditional sectors face substantial challenges in efficiently managing their production capacity as well as distribution channels to meet the service requirements of a dynamic market and reap the highest possible profit.

As a key determinant of the service level, equipment capacity management is critical to the success of many firms. For manufacturers in such capital-intensive industries as the semiconductor, electronic, automotive, and pharmaceutical industries, equipment capacity expenditures constitute about one quarter of the total revenue and roughly two thirds of the total manufacturing costs (Hertzler 2009). Intel, for instance, spent on average 5.3 billion USD annually on capital equipment purchase and development from 2004 to 2009 (Intel Corporation 2009). On top of the substantial costs, capacity decisions faced by these industries, especially the semiconductor and electronic industries, are very difficult to make due to several reasons. First, the consumer market is highly volatile. The Semiconductor Industry Association once reported that global semiconductor sales in the first quarter of 2009 declined 29.9% from the first quarter of 2008 due to the sweeping economic recession. However, sales rose

47.6% in May 2010 from May 2009 when the overall economy started improving again (Detar 2010). In recent years, the fast growth of mobile and cloud computing has also started shifting the original market layout and further exacerbates the demand randomness in related industries. This high level of market uncertainty together with a very long supply leadtime of capital equipment (quarters or even years) sharply increases the difficulty of accurate demand forecasting and capacity planning. In addition, stockouts can be very costly. A single computer chip or consumer electronic device is usually sold at hundreds of dollars, and the indirect loss of goodwill is even more damaging to firms' long-term revenue. While this fact certainly underscores the importance of avoiding shortages, it does not mean firms should always err on the side of having idle capacity, since capacity is cumulative and any over-investment is not only expensive but also irreversible. Therefore, the ability to respond rapidly to the changing market demand while minimizing unnecessary capital expenditures is both important and challenging to firms in capital-intensive industries.

On the selling side, as market dynamics become more complex, it is no longer sufficient for a business to focus solely on its main distribution channel and overlook opportunities for additional outlets for its product and service. Firms are identifying increasingly diversified market platforms where they can sell products to different groups of buyers at different times and prices, and it is important for them to leverage multiple distribution channels to enlarge their market reach and hedge against potential demand and price risks. In fact, more than 80% of US retailers were already offering multi-channel transactional capabilities in 2007 (Lovett and Anand 2007) and this number is even higher today. While for traditional retailers a channel mix usually involves conventional bricks-and-mortar stores, online e-commerce, and catalog sales, this is not the only type of channel choice that firms may encounter. Distribution channels may sometimes refer to the different contract formats in which firms interact with their customers. For instance, commodity suppliers such as Rio Tinto face the distribution choice between a fixed-price contract channel and a real-time spot market (Lee 2007); hotel chains and rental car firms can distribute their capacity either through their own websites or anonymous discount aggregators such as Priceline.com. For all these businesses it is crucial to determine the optimal split of sales volumes among different distribution channels to maximize their potential profits.

## 1.2. Research Overview

In this dissertation, we address the aforementioned challenges by investigating how firms can operate multiple channels to build their production capacity or distribute sales volumes to meet the needs of an ever-changing market. It consists of two major parts: In the first part (Chapters 2 and 3), we focus on the sourcing side and study how firms in capital-intensive industries can efficiently purchase their production equipment through two different supply channels with complementary leadtimes and prices. In the second part (Chapter 4), we switch to the selling side and investigate how a commodity supplier can utilize both a fixed-price contract channel and a spot channel to optimally distribute his sales volumes. Below are some more details.

In Chapter 2, we study a finite-horizon, periodic-review, dual-source equipment capacity expansion problem with demand backlogging. In each period, the manufacturer acquires updated demand information and procures manufacturing equipment from two suppliers with different leadtimes and prices: a *base* supplier which is inexpensive but slow and a *flexible* supplier which is fast but expensive. The combination of these two sources provides the firm the flexibility to mitigate demand risk at a potentially lower cost. We prove that even in the simplest setting, i.e., when the leadtimes are consecutive and zero-one, only the flexible orders follow a state-dependent base-stock policy. The base orders, on the other hand, follow only a partial-base-stock policy; the expand-to capacity position for the base source could be decreasing in the initial capacity level due to the effect of backlogging. We thus conclude that the dynamic dual-source capacity expansion problem with backorders is inherently different and more complex than its inventory counterpart as well as the dynamic dual-source capacity expansion problem with lost sales, and we call for solutions to this rather complex problem.

In Chapter 3, we propose a dual-mode equipment procurement (DMEP) framework as a direct response to the challenge posed in Chapter 2. The DMEP framework combines the same dual-channel procurement strategy with capacity options and a detailed forecast revision mechanism. It consists of three layers: a strategic *contract negotiation* layer, in which the firm chooses the best combination of leadtime and price for each supply mode from the contract menu, a tactical *reservation* layer, in which the firm reserves total equipment procurement quantities from the two supply modes by paying the supplier a reservation fee up front before the planning horizon starts, and an operational *execution* layer, in which the firm acquires the latest demand information in each period and orders equipment from the two supply modes following a rolling-horizon algorithm. With a comprehensive numerical analysis, we quantify the value of the added flexibility of dual-mode equipment procurement for the firm. Implementation of this approach has leveraged model structure and details to provide the types of sensitivity analysis needed by decision-makers to understand and take advantage of the subtleties of this improvement in the capital equipment acquisition process. The annual savings on capital procurement at Intel due to implementing DMEP are estimated to exceed tens of millions of dollars.

In Chapter 4, we study a commodity-selling problem in which a large commodity supplier strategically allocates his production capacity between a fixed-price contract channel and a spot channel to maximize his total sales income. We discuss two settings: one in which the equilibrium spot price follows an exogenous random distribution (open spot market) and one in which the equilibrium spot price is endogenously determined by the spot demand curve and the spot supply curve, both of which can be affected by the supplier's capacity allocation decision (closed spot market). We identify the supplier's optimal allocation policies under both circumstances and demonstrate how the optimal decisions change with the key model parameters. For the open spot market setting, we find that the demand-price correlation and a risk-averse attitude are two reasons for the supplier to adopt a dual-channel strategy. The supplier should allocate more quantity to the spot channel if the contract channel demand and the spot price are more positively correlated, and he should allocate more to the contract channel if he is more risk-averse. For the closed spot market setting which further includes a contract trading stage and a spot trading stage, we show that a dual-channel policy is optimal in the first stage if the shifting effect of the supplier's spot allocation quantity on the default supply curve is stronger than the shifting effect of the unfulfilled contract channel demand on the default demand curve. Further, we demonstrate that it is not necessarily optimal to sell all leftover quantities in the spot market during the second stage. We also quantify the average improvement in profit of adopting a dual-channel strategy versus using a single contract channel or a single spot channel through numerical analysis. Although this research is motivated by the business practice in the commodity industry, the modeling methodology and managerial insights are certainly applicable to other industries with similar channel choices to make.

The layout of the dissertation is as follows: we discuss the related literature, present the details of each model, and demonstrate the corresponding results in Chapters 2 through 4. We conclude the dissertation in Chapter 5, summarizing the major contributions and indicating future research directions. All the proofs are provided in the appendix.

# Chapter 2

# Dynamic Dual-Source Capacity Expansion Problem with Backorders

## 2.1. Introduction

From 2004 to 2009, as a leader in the semiconductor industry, Intel has spent on average 5.3 billion USD annually on capital equipment purchase and development (Intel Corporation 2009), an amount greater than the GDP of 51 countries in the world in 2009 (World Bank 2009). In fact, in capital-intensive industries (such as the semiconductor, electronic, automotive, and pharmaceutical industries), capital expenditures often occupy 20-30% of the total revenue, up to 50% of the company's gross margin, and about two-thirds of the manufacturing costs (Hertzler 2009). While the right volume of capital equipment guarantees the smoothness and continuity of the production process, the inappropriate purchase of this equipment may lead to either idle capacity or unsatisfied demand, both of which can result in a considerable loss of profit. Thus, choosing the right amount of equipment to purchase is a vital decision for managers in these industries.

Unfortunately, this is not a straightforward decision. Due to rapid technological change and fierce market competition, demand faced by these industries, in particular the semiconductor and electronics industries, is highly uncertain. Purchasing the proper amount of manufacturing equipment requires an accurate demand forecast, which is difficult to obtain. Additionally, the leadtime (including production, transportation, and installation) of manufacturing equipment can be as long as months, or even years, which increases the difficulty of demand forecasting, and hence aggravates the risk of advance purchase. Moreover, unlike excess inventory of products, which may still be used to satisfy future demand, over-investment in manufacturing equipment is irreversible and idle equipment may continue to be idle in the future even if no additional equipment is purchased. This cumulative property alone implies that classical inventory control strategies may not be applicable to managing equipment capacity.

This chapter aims to study the aforementioned problem. We take the perspective of a manufacturer in a capital-intensive industry and analyze a periodic-review, finite-horizon, dual-source capacity expansion problem with demand forecast updates (*capacity* here refers to the capacity of manufacturing equipment). Dual-sourcing strategies are commonly adopted in inventory management for mitigating supply costs and risks; here, we introduce such strategies to the capacity expansion environment. We suppose the manufacturer is starting mass production of a given product and has to build capacity over time as demand for his product increases and better market information becomes available. He can order capacity from two suppliers, a *base* supplier and a *flexible* supplier, with different leadtimes and prices. At each decision period, the manufacturer chooses the amount of capacity to order from each supplier given the latest demand information, with the goal of maximizing his expected total profit-to-go. Unmet demand is fully backlogged.

## 2.2. Literature Review

The dual-sourcing problem has been studied in the context of inventory since the early 1960s. The extant literature establishes that for the case of consecutive leadtimes, a two-level (modified) base-stock policy is optimal (Daniel 1963, Fukuda 1964, Whittemore and Saunders 1977, Yazlali and Erhun 2007). Yan *et al.* (2003) and Sethi *et al.* (2001) investigate this problem under forecast updates and show that a modified base-stock policy continues to be optimal. For the more general nonconsecutive leadtimes case, the optimal policy can be quite complex and lacks structure (Scheller-Wolf and Tayur 1998, Feng *et al.* 2006, Yazlali and Erhun 2009).

Despite the existence of literature on the dual-sourcing inventory problem, research on the dual-sourcing capacity problem is rare. Capacity behaves very differently from inventory in that it can be repeatedly used from one period to another; hence classical inventory management policies may not be applicable to capacity planning problems. Capacity decisions have been extensively analyzed under single-sourcing (see Van Mieghem 2003 for a review of this literature). However, to the best of our knowledge, Chao *et al.* (2009) and Peng *et al.* (2010) are the only papers that study the dual-sourcing capacity problem. Chao *et al.* (2009) investigate a dual-source setting with consecutive leadtimes where demand in excess of capacity is lost; they establish that the optimal capacity expansion policy with the fast (flexible) source is base-stock. The authors further demonstrate that when the capacity obsolescence rate is deterministic, the optimal policy for capacity expansion through the slow (base) source is also base-stock. We study a problem similar to Chao *et al.*'s deterministic-capacity-obsolescence-rate setting. However, our model differs from theirs in two ways: *(i)* the demand in excess of capacity is backlogged instead of lost, and *(ii)* we incorporate a forecast updating process. Backlogging is a more realistic assumption for many industries, especially in manufacturing, but it also immediately escalates the complexity of the problem by adding an extra state variable to the dynamic programming formulation. As a result, we show that a base-stock policy is no longer optimal for the base source. Forecast updating allows us to investigate the value of information, which is a critical component due to extremely long leadtimes associated with capacity investment decisions. Interested readers may refer to Peng *et al.* (2010) (or Chapter 3 of this dissertation) for a heuristic solution to a general dual-source capacity expansion problem.

The rest of this chapter is arranged as follows: in Sections 2.3 and 2.4, we construct the model and derive the analytical results. We show that, even in the simplest setting possible (i.e., when leadtimes are consecutive and zero-one) the capacity expansion problem is categorically different than its inventory counterpart and the irreversibility of the capacity investment decisions complicates the analysis significantly. Section 2.5

provides additional managerial insights through numerical analysis. Section **??** briefly discusses an extension where inventory carry-over is allowed. Section 2.6 concludes this chapter. Proofs of all results are in the appendix.

## 2.3. Model Definition

We consider a finite-horizon, periodic-review, dual-source capacity expansion model with demand backlogging and forecast updates. A manufacturer needs an efficient equipment procurement strategy to optimally match his production capacity with growing but fluctuating demand over time. During each period $n$, $n = 0, 1, \cdots, N$, he orders capacity from two suppliers: an inexpensive-but-slow *base* supplier and a fast-but-expensive *flexible* supplier. He does so to maximize his total expected profit over the entire planning horizon. We purposely analyze the simplest setting; i.e., we assume consecutive leadtimes. Furthermore, the base supplier's leadtime is one period, thus the flexible supplier delivers instantaneously. The unit price for the base supplier is $c_b$ and for the flexible supplier is $c_f > c_b$.

Before introducing the details of our model, we first explain the underlying market demand process. Demand $D_n$ in each period $n$ is composed of three parts: a deterministic component $\mu_n$, the initial market information $\varepsilon_n^1$, and the final market information $\varepsilon_n^2$. The manufacturer knows $\mu_n$, and observes $\varepsilon_n^1$ at the beginning of period $n$ and $\varepsilon_n^2$ at the end of period $n$. Both $\varepsilon_n^1$ and $\varepsilon_n^2$ are random variables. Market information for different periods and different market information in the same period are all assumed to be independent of each other. Given the above assumptions, $D_n$ can be expressed as $D_n = g(\varepsilon_n^1, \varepsilon_n^2, \mu_n)$ where $g(\cdot)$ is any Borel-measurable function.

This simple demand structure fits into the classical *martingale model of forecast evolution* (MMFE) (Hausman 1969, Heath and Jackson 1994, Graves *et al.* 1986). Two commonly used functions are the additive form $D_n = \varepsilon_n^1 + \varepsilon_n^2 + \mu_n$ and the multiplicative form $D_n = \varepsilon_n^1 \varepsilon_n^2 \mu_n$. This demand model allows us to easily capture demand forecast updates ; we do so in the manner of Sethi *et al.* (2001).

The manufacturer determines the optimal amount of equipment to order from both suppliers at each period $n$. The sequence of events is as follows (Figure 2.1): ($i$) At the beginning of period $n$, the manufacturer observes the current capacity position

$x_n + B_{n-1}$ where $x_n$ is his on-hand capacity and $B_{n-1}$ is the on-order capacity from the base supplier in period $n-1$. He also observes the initial market information $\varepsilon_n^1$ and any backlogged demand $y_n$ from the previous period. $(ii)$ The manufacturer places a flexible order $F_n$ at unit price $c_f$ and a base order $B_n$ at unit price $c_b$ from the two suppliers, respectively. Note that in period 0 only base orders are placed since demand will not materialize until period 1; and in period $N$ only flexible orders are placed since this is the last period in the selling horizon. $(iii)$ Orders $F_n$ and $B_{n-1}$ arrive. $(iv)$ The final market information $\varepsilon_n^2$ is revealed and demand $D_n$ is realized. $(v)$ Given the on-hand equipment capacity $x_n + B_{n-1} + F_n$, production is carried out to satisfy demand at a unit profit margin of $p_n$. Without loss of generality, we assume each unit of capacity can be used to process only one product every period. $(vi)$ Any unsatisfied customer demand $y_{n+1}$ is backlogged. Note that when base orders are placed for period $n$ $(B_{n-1})$, the buyer still faces a lot of demand uncertainty ($\varepsilon_{n-1}^2$, $\varepsilon_n^1$, and $\varepsilon_n^2$). However, he has much improved demand information (both in terms of demand realization $\varepsilon_{n-1}^2$ and an updated forecast $\varepsilon_n^1$) with the flexible source $F_n$. As such, our demand structure does not limit the value of flexible source to demand realizations, but also captures forecast updates. That is, when a large demand forecast update $\varepsilon_n^1$ is observed, the flexible source can be used to dampen the risk of capacity underage.

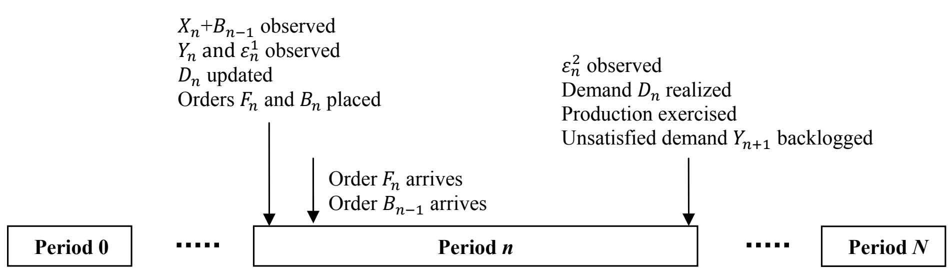



Figure 2.1: Sequence of Events in the Planning Horizon

There are several cost and revenue parameters that affect the manufacturer's decisions. We assume that the manufacturer's unit profit margin $p_n$ is decreasing in $n$, hence it is more profitable to satisfy demand earlier rather than later. Since we allow for backorders, the decreasing margin also functions as the backlog penalty. On-hand capacity incurs unit holding/maintenance cost $c_h$ per period. There is zero salvage value for on-hand capacity after the horizon ends. This can be justified by the fact that leading manufacturers do not cheap-sell their idle capacity for fear of revealing crucial technology to competitors. Any unsatisfied demand after the terminal period $N$ incurs an additional unit penalty $c_u$, which may be due to the fact that an expensive alternative source is used to satisfy this demand. As our goal is to understand the optimal tactical level capacity decisions, we suppress the executional level inventory problem and assume that production in a period will not exceed demand. We provide two justifications for this assumption: 1. Many firms (such as Dell Inc.) in high-tech industry consider it especially risky to carry inventory and struggle to keep their stock level as low as possible; 2. Even for those that do maintain a constant safety stock level, production would essentially be built-to-order had that constant level been removed. However, we do have a brief discussion about an inventory carry-over scenario in Section A.1. We assume that all random variables have finite mean and variance, and we impose a discount factor $\delta$, $0 < \delta < 1$, per period.

The sequential decision problem of choosing the optimal $B_n$ and $F_n$ for all $n$ can be formulated as a dynamic programming model. To simplify the notation, we introduce a new state variable $\tilde{x}_n$ as the capacity position at the beginning of period $n$, which is defined as $\tilde{x}_n = x_n + B_{n-1}$. The other state variables are $y_n$, the unmet demand from last period, and $\epsilon_n^1$, the observed initial market information at period $n$ ($\varepsilon_n^1 = \epsilon_n^1$). We construct the $(N+1)$-stage dynamic programming model as follows ($n = 0, 1, \cdots, N$):

$$\begin{aligned} J_n(\tilde{x}_n, y_n, \epsilon_n^1) \;=\; \max_{B_n, F_n} \mathbb{E}[&p_n \min\{g_n(\epsilon_n^1, \varepsilon_n^2, \mu_n) + y_n, \tilde{x}_n + F_n\} - c_b B_n - c_f F_n \\ &- c_h(\tilde{x}_n + F_n) + \delta J_{n+1}(\tilde{x}_{n+1}, y_{n+1}, \varepsilon_{n+1}^1)] \qquad (2.3.1) \end{aligned}$$

subject to $B_n \geq 0$, $F_n \geq 0$, and $J_{N+1}(\cdot) = -c_u y_{N+1}$. In the objective function of (2.3.1), the first term denotes the profit; the second and the third terms are the base and flexible capacity ordering costs; the fourth term captures the holding cost of the on-hand capacity; and the last term is the discounted profit-to-go. In the final period $N$, the manufacturer incurs a penalty cost for any unsatisfied demand at the end of the planning horizon. The states are updated according to the following equations: capacity position $\tilde{x}_{n+1} = \tilde{x}_n + B_n + F_n$ where $\tilde{x}_0 = 0$; unmet demand $y_{n+1} = (y_n + g_n(\epsilon_n^1, \varepsilon_n^2, \mu_n) - (\tilde{x}_n + F_n))^+$ where $y_1 = 0$ and we use the convention $(x)^+$ for $\max\{x, 0\}$.

Next, we analyze this dynamic program to define a procurement strategy for the manufacturer.

## 2.4. Analytical Results

We have formulated a dual-source capacity expansion model with demand forecast updates. In order to analyze the dynamic program, we must first structure the forecast update process. To that end, we assume that demand $D_n$ is in additive form, that is,

$$D_n = g_n(\varepsilon_n^1, \varepsilon_n^2, \mu_n) = \varepsilon_n^1 + \varepsilon_n^2 + \mu_n.$$

The additive form is commonly used in the literature; e.g., Gallego and Özer (2001), Chen and Lee (2009). With this form, we can further simplify the model by combining the unmet demand $y_n$ with the initial market information $\varepsilon_n^1$, both of which are revealed before the ordering decision is made. We denote this new term $\tilde{y}_n = y_n + \varepsilon_n^1$ as the *modified* backorder level. We also update the decision variables and work with expand-to capacity positions, $\tilde{x}'_n = \tilde{x}_n + F_n$ and $\tilde{x}''_n = \tilde{x}'_n + B_n$. As a result, the original model can be re-written as follows:

$$J_n(\tilde{x}_n, \tilde{y}_n) = \max_{\tilde{x}_n \le \tilde{x}'_n \le \tilde{x}''_n} V_n(\tilde{x}_n, \tilde{y}_n, \tilde{x}'_n, \tilde{x}''_n),$$

where

$$V_n(\tilde{x}_n, \tilde{y}_n, \tilde{x}'_n, \tilde{x}''_n) = \mathbb{E}\Big[p_n \min\left\{\tilde{y}_n + \varepsilon_n^2 + \mu_n, \tilde{x}'_n\right\} - c_b(\tilde{x}''_n - \tilde{x}'_n) - c_f(\tilde{x}'_n - \tilde{x}_n) - c_h\tilde{x}'_n + \delta J_{n+1}(\tilde{x}''_n, y_{n+1} + \varepsilon_{n+1}^1)\Big] \quad (2.4.1)$$

for $n = 0, 1, \cdots, N$; $y_{n+1} = (\tilde{y}_n + \varepsilon_n^2 + \mu_n - \tilde{x}'_n)^+$; and $J_{N+1}(\cdot) = -c_u y_{N+1}$.

To determine the optimal capacity expansion policy for the above model, we first need to establish some structural properties of the objective function in (2.4.1).

**Lemma 2.4.1.** *$J_n(\tilde{x}_n, \tilde{y}_n) = \tilde{J}_n(\tilde{x}_n, \tilde{y}_n) + G_n(\tilde{y}_n)$ where*

$$
\begin{aligned}
G_n(\tilde{y}_n) &= p_n \tilde{y}_n + \mathbb{E}\left[\sum_{k=n+1}^{N} \delta^{k-n} p_k \varepsilon_k^1\right] + \mathbb{E}\left[\sum_{k=n}^{N} \delta^{k-n} p_k (\varepsilon_k^2 + \mu_k)\right], \\
\tilde{J}_n(\tilde{x}_n, \tilde{y}_n) &= \max_{\tilde{x}_n \le \tilde{x}_n' \le \tilde{x}_n''} \mathbb{E}\Big[ - (p_n - \delta p_{n+1})(\tilde{y}_n + \varepsilon_n^2 + \mu_n - \tilde{x}_n')^+ - c_b(\tilde{x}_n'' - \tilde{x}_n') \\
&\qquad - c_f(\tilde{x}_n' - \tilde{x}_n) - c_h \tilde{x}_n' + \delta \tilde{J}_{n+1}(\tilde{x}_n'', y_{n+1} + \varepsilon_{n+1}^1)\Big], \qquad (2.4.2)
\end{aligned}
$$

*for $n = 0, 1, \cdots, N$; $y_{n+1} = (\tilde{y}_n + \varepsilon_n^2 + \mu_n - \tilde{x}_n')^+$; $p_{N+1} = 0$ and $\tilde{J}_{N+1}(\cdot) = -c_u y_{N+1}$.*

By Lemma 2.4.1, $J_n(\tilde{x}_n, \tilde{y}_n)$ is the summation of $\tilde{J}_n(\tilde{x}_n, \tilde{y}_n)$ and a linear function of $\tilde{y}_n$. This transformation is useful since the second term is independent of $\tilde{x}_n$ and hence does not affect the capacity decision. Therefore, the two stochastic decision problems defined by $J_n(\cdot,\cdot)$ and $\tilde{J}_n(\cdot,\cdot)$ should (potentially) obey the same concavity structure and have the same optimal solution.

**Lemma 2.4.2.** *For all $n$, $\tilde{J}_n(\tilde{x}_n, \tilde{y}_n)$ is decreasing in $\tilde{y}_n$ and concave in $(\tilde{x}_n, \tilde{y}_n)$. Also, the objective function in Equation (2.4.2) is concave in $(\tilde{x}_n, \tilde{y}_n, \tilde{x}_n', \tilde{x}_n'')$.*

**Proposition 2.4.3.** *$J_n$ is concave in $(\tilde{x}_n, \tilde{y}_n)$; the objective function $V_n$ is concave in $(\tilde{x}_n', \tilde{x}_n'')$ for any given $\tilde{x}_n$ and $\tilde{y}_n$.*

The joint concavity of the objective function of the dual-sourcing model in $\tilde{x}_n$ and $\tilde{y}_n$ enables us to derive our next result.

**Proposition 2.4.4.** *For the flexible source, a state-dependent base-stock policy is optimal.*

Proposition 2.4.4 demonstrates that a well-behaved structural policy exists for the flexible source. However, the base-stock policy is not necessarily optimal for the base source. Defining $S_n^F(\tilde{y}_n)$ as the optimal base-stock level for the flexible source and $S_n^B(\tilde{y}_n) \ge S_n^F(\tilde{y}_n)$ as the optimal partial-base-stock level for the base source, we can prove the optimality of a partial-base-stock policy for the base source.

**Proposition 2.4.5.** *For the base source, there exists a partial-base-stock policy with parameter $\underline{S}_n^B(\tilde{y}_n)$ satisfying $\underline{S}_n^B(\tilde{y}_n) \ge S_n^F(\tilde{y}_n)$ such that:*

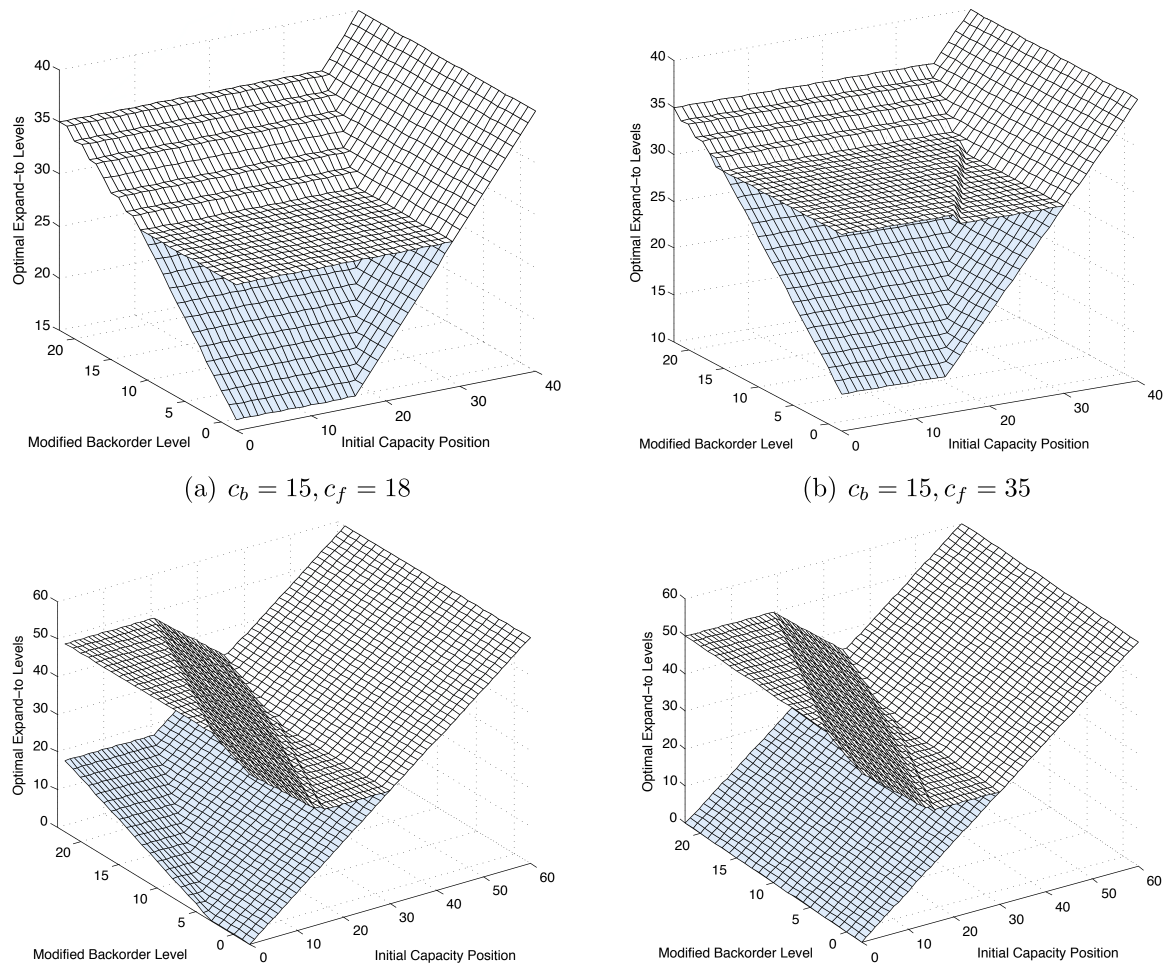
*(i) if* $\tilde{x}_n \leq S_n^F(\tilde{y}_n)$*, it is optimal to expand the capacity position to* $S_n^B(\tilde{y}_n)$*;*

*(ii) if* $S_n^F(\tilde{y}_n) < \tilde{x}_n \leq \underline{S}_n^B(\tilde{y}_n)$*, the optimal expand-to capacity position depends on* $\tilde{x}_n$ *and* $\tilde{y}_n$*;*

*(iii) if* $\tilde{x}_n > \underline{S}_n^B(\tilde{y}_n)$*, then it is optimal not to order from the base source.*

Figure 2.2 graphically demonstrates how the optimal expand-to capacity levels of the base and flexible sources change with the cost parameters as well as with the initial capacity position. The figure displays results for $n = 2$, but the insights derived are applicable to all periods. In each of the four graphs, the $x$-axis denotes the initial capacity position $\tilde{x}_2$; the $y$-axis denotes the modified backorder level $\tilde{y}_2$; and the vertical axis represents the optimal expand-to capacity positions for both supply sources with respect to each state $(\tilde{x}_2, \tilde{y}_2)$. We fix the base ordering cost $c_b = 15$ in all cases and change the flexible ordering cost $c_f$. In Figure 2.2(a), we set $c_f = 18$ and observe that $S_2^F(0) = 18$, $S_2^B(0) = \underline{S}_2^B(0) = 29$. This is a special case where a state-dependent base-stock policy is optimal for both the flexible and base sources. Sufficient flexible capacity is ordered due to the relatively low flexible cost, hence (almost) no demand will ever be backlogged into the next period, which leads to the optimality of a base-stock policy for the base source. By gradually increasing the flexible ordering cost $c_f$ in Figures 2.2(b)-2.2(d), we observe that for medium values of initial capacity position $\tilde{x}_2$, the state-dependent base-stock policy fails. The more we increase the gap between the unit prices, the more this failure becomes prominent. Furthermore, the optimal expand-to capacity position for the base source may actually decrease in the initial capacity position. When the flexible ordering cost is high, the consequently insufficient amount of flexible orders may cause backorders; nevertheless, the higher the initial capacity level, the more of these potential backorders can be eliminated in the current period, thus decreasing the required capacity position for the next period.

The failure of the optimality of a state-dependent base-stock policy for the base source is a stark departure from the related inventory literature and is counterintuitive at first. In the capacity expansion setting, the cumulative capacity automatically serves as an additional source. The existence of this source is not consequential when there are no backorders as in Chao *et al.* (2009). However, if there are backorders,




(a) $c_b = 15, c_f = 18$ (b) $c_b = 15, c_f = 35$

(c) $c_b = 15, c_f = 45$ (d) $c_b = 15, c_f = 65$

Figure 2.2: Optimal Expand-to Levels under Different Flexible Ordering Costs for $n = 2$ ($N = 3$, $\vec{p} = [p_1, p_2, p_3] = [60, 50, 40]$, $c_h = 4$, $c_u = 10$, $\mu = [\mu_1, \mu_2, \mu_3] = [10, 17, 30]$, $\varepsilon_n^1$ and $\varepsilon_n^2$ for every period $n$ are independent and identically distributed and satisfy a discrete uniform distribution on $[-2, 2]$, and $\delta = 0.98$)

then this source acts as a fictitious delivery mode for which the manufacturer even does not have to place an order. As such, it can be thought as a delivery mode with a leadtime of $-1$. Therefore, even a dual-source problem with consecutive zero-one leadtimes acts as a multi-source problem. With this interpretation, our result is consistent with Feng *et al.* (2006), where the authors show that only the fastest two modes (in our case the fictitious mode and the flexible source which is consecutive with it) have a base-stock policy. Note that the modified backorder level by definition includes the initial market information ($\epsilon_n^1$). Thus, the initial market information can also contribute to the observation above. Having said that, even when we eliminate this additional information, the results and observations remain unchanged.

Even if there are certain ranges of $\tilde{x}_n$ in which an optimal base-stock policy exists (as in Figure 2.2(b)), the base-stock level $S_n^F(\tilde{y}_n)$ for the flexible source and the partial-base-stock level $S_n^B(\tilde{y}_n)$ for the base source are not trivial to calculate. We first observe from the figure that both $S_n^F(\tilde{y}_n)$ and $S_n^B(\tilde{y}_n)$ are weakly increasing in $\tilde{y}_n$; furthermore, our next proposition establishes certain linear relations that can further simplify the derivation of $S_n^F(\tilde{y}_n)$ and $S_n^B(\tilde{y}_n)$:

**Proposition 2.4.6.** *The base-stock level $S_n^F(\tilde{y}_n)$ for the flexible source and the partial-base-stock level $S_n^B(\tilde{y}_n)$ for the base source satisfy $S_n^F(\tilde{y}_n) = S_n^F(0) + \tilde{y}_n$, $S_n^B(\tilde{y}_n) = S_n^B(0)$ if $\tilde{y}_n \leq S_n^B(0) - S_n^F(0)$, $n = 1, 2, \cdots, N-1$; and $S_N^F(\tilde{y}_N) = S_N^F(0) + \tilde{y}_N$.*

Proposition 2.4.6 states that when the modified backorder level $\tilde{y}_n$ is lower than a certain value ($\tilde{y}_n \leq S_n^B(0) - S_n^F(0)$), the flexible source base-stock level $S_n^F(\tilde{y}_n)$ is simply the summation of $\tilde{y}_n$ and a constant $S_n^F(0)$ (hence is linearly increasing in $\tilde{y}_n$), and that the base source partial-base-stock level $S_n^B(\tilde{y}_n)$ remains constant. We can observe this fact by revisiting Figure 2.2(b). In accordance with Proposition 2.4.6, $S_n^F(\tilde{y}_n)$ increases linearly and $S_n^B(\tilde{y}_n)$ stays constant as $\tilde{y}_n$ increases until the two levels coincide, after which point they increase together nonlinearly.

In Proposition 2.4.5, we only claim that for a general $N$-period problem, the optimal base expand-to position depends on the initial capacity level; Figure 2.4.5 further demonstrates that there are cases where the base expand-to position decreases in the initial capacity level. Proposition 2.4.7 below formalizes this observation for a two-period case.

**Proposition 2.4.7.** *For a special case with two periods only, i.e., $N = 2$, we have:*

*(i) $\tilde{J}_2(\tilde{x}_2, \tilde{y}_2)$ is supermodular;*

*(ii) The first period objective function[1] is submodular in $(\tilde{x}_1', \tilde{x}_1'')$;*

[1] Refer to the proof for a rearrangement of the objective function.

*(iii) For $S_1^F(\tilde{y}_1) < \tilde{x}_1 \leq \underline{S}_1^B(\tilde{y}_1)$, the optimal base expand-to capacity position $\tilde{x}_1''$ decreases in $\tilde{x}_1$.*

Before concluding this section, we make an observation about the single-source capacity expansion problem with demand backlogging. Interestingly, for even this case, the optimal strategy may lack structure:

**Corollary 2.4.8.** *A state-dependent base-stock policy is optimal for the single-source capacity expansion problem with a leadtime $L = 0$ and demand backlogging. However, this result cannot be extended for a leadtime $L > 0$.*

Again, this result is quite counterintuitive, since in the dynamic inventory control problem with demand backlogging, we can easily convert the positive leadtime case to the zero leadtime case by defining *inventory position* (on-hand inventory plus on-order inventory) as the new state variable and using *leadtime demand* to replace the original single-period demand. By doing so, one can reestablish the optimality of the base-stock policy. Unfortunately, this method ceases to apply in the capacity expansion setting. In fact, for the single-source capacity expansion problem with demand backlogging and a leadtime of $L$ ($> 0$) periods, there need to be $L + 2$ state variables in the dynamic programming formulation: one state for the backlogged demand from previous period; one state for the on-hand capacity at the beginning of the current period; and $L$ states for the on-order capacity that is still in the pipeline and will arrive within the next $L$ periods. The reason is that due to the cumulative property of tool capacity, in order to keep track of the backlogged demand for a certain period, the backlogged demand at each of the previous periods also has to be recorded. This implies that one has to know the precise capacity level for every period within the leadtime duration. With $L+1$ state variables denoting the capacity positions for different periods, no base-stock type policy would exist.

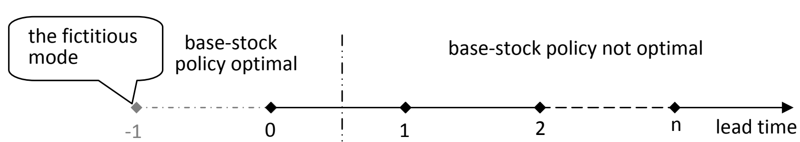



Figure 2.3: The Optimality of Base-Stock Policy in Single-Source Capacity Expansion

There is another intuitive explanation for why a base-stock policy is not optimal for the case of positive leadtime. In the capacity expansion setting, the cumulated capacity so far automatically serves as a fictitious delivery mode with leadtime $-1$—the order arrives even before it is placed. Therefore, a single-source capacity expansion problem can always be interpreted as a multi-source inventory problem with one $-1$-leadtime delivery mode. As we have already introduced, Feng et al. (2006) show that for multi-source consecutive-leadtime inventory control problems, only the fastest two modes have base-stock policies. So, if the single source has a zero leadtime, which is consecutive with $-1$, then a base-stock policy should be optimal. If however the single source has a positive ($L > 0$) leadtime, then we can insert several imaginary supply modes (with leadtimes $0, 1, \cdots, L-1$) to transform the problem into one with consecutive modes by setting the cost for any imaginary mode to be sufficiently high (so that no optimal policy would issue orders to the imaginary modes). And Feng et al. (2006) now tell us that a base stock policy is generally not optimal for this single source. The explanation is summarized in Figure 2.3.

## 2.5. Managerial Insights

In the above section we presented some important analytical properties of the dual-source capacity expansion problem. In this section, we generate additional managerial insights through numerical analysis. Specifically, we examine the performance of a two-level-base-stock heuristic, the value of having dual capacity sources, and the value of having demand forecast updates. The benchmark parameter values for all experiments are set to the same as in Figure 2.2.

### 2.5.1 Two-Level Base-Stock Policy

We notice that when $c_f = 35$ in Figure 2.2(b), both the width of the middle region and the decreasing trend are mild, and the ordering policy for the base source is approximately base-stock – $\bar{S}_n^B(0)(=31)$ and $\underline{S}_n^B(0)(=30)$ are very close to each other. Table 2.1 below shows that under this circumstance by following a two-level base-stock policy with base-stock levels 16 and 31, instead of following the optimal ordering policy, the manufacturer sacrifices only a small portion of profit. Similar results are observed as we vary the value of $c_f$ between 15 and 35.

Table 2.1: Optimal Strategy (OS) vs. Two-Level Base-Stock Policy (TLBS) Profit Comparison ($\times 10^3$; $c_b = 15$; $c_f = 35$; $n = 2$; $\tilde{y}_2 = 0$; % $\downarrow$ means percentage profit decrease)

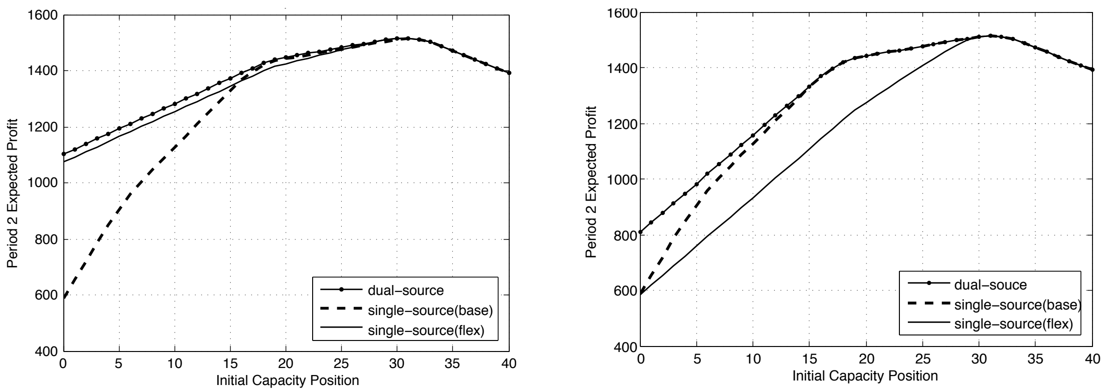
| init capacity | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| OS | 1.420 | 1.435 | 1.442 | 1.449 | 1.456 | 1.463 | 1.470 | 1.477 | 1.484 | 1.491 | 1.498 |
| TLBS | 1.418 | 1.431 | 1.438 | 1.445 | 1.452 | 1.459 | 1.466 | 1.473 | 1.480 | 1.487 | 1.494 |
| (% $\downarrow$) | 0.14% | 0.28% | 0.28% | 0.28% | 0.27% | 0.27% | 0.27% | 0.27% | 0.27% | 0.27% | 0.27% |

**Remark.** When faced with moderate flexible ordering cost, a capacity manager can actually follow a two-level state-dependent base-stock policy (characterized by the base-stock level $S_n^F(\tilde{y}_n)$ for flexible source and the base-stock level $\bar{S}_n^B(\tilde{y}_n)$ for base source) when ordering from the two sources. This will largely simplify the decision making process while compromising only a small amount of profit. Meanwhile, if the demand in each period is increasing stochastically at a fast enough rate so that the partial-base-stock level for the base source during period $n$ is lower than the base-stock level for the flexible source in period $n+1$, then on the optimal path the firm will always be operating within the well-behaved region where a two-level base-stock policy is optimal.

### 2.5.2 The Value of Dual-Sourcing

**Proposition 2.5.1.** *The optimal expected profit achieved under dual-source capacity expansion is higher than the optimal profit achieved under single-source capacity expansion, irrelevant to the price/cost structure and regardless of whether the single-source is using the base supplier or the flexible supplier.*

Proposition 2.5.1 states that dual-sourcing enables the manufacturer to reap a higher profit than single-sourcing, either with the base supplier or the flexible supplier. In practice, however, manufacturers may not be willing to sign contracts with two capacity suppliers (or with the same supplier for different supply modes) unless the gain from dual-sourcing is nontrivial, since contracting and supplier management themselves involve administrative costs. We demonstrate below that the additional benefit of having dual-source can be significant.




(a) $c_b = 15,\ c_f = 18$ (b) $c_b = 15,\ c_f = 35$

Figure 2.4: Single-Source v.s. Dual-Source Optimal Profit ($n = 2,\ \tilde{y}_2 = 0$)

Table 2.2: Single-Source vs. Dual-Source Profit Comparison ($\times 10^3$; $n = 2$; $\tilde{y}_2 = 0$; % ↓ means percentage profit decrease compared with the dual-source case)

| initial capacity | | 0 | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | dual-source | 1.103 | 1.139 | 1.175 | 1.211 | 1.247 | 1.283 | 1.319 | 1.355 | 1.391 | 1.427 | 1.448 |
| low | flex-source | 1.075 | 1.111 | 1.147 | 1.183 | 1.219 | 1.255 | 1.291 | 1.327 | 1.363 | 1.399 | 1.425 |
| $c_f$ | (% ↓) | 2.54% | 2.46% | 2.38% | 2.31% | 2.25% | 2.18% | 2.12% | 2.07% | 2.01% | 1.96% | 1.59% |
| | base-source | 0.587 | 0.720 | 0.848 | 0.960 | 1.046 | 1.127 | 1.209 | 1.290 | 1.367 | 1.420 | 1.442 |
| | (% ↓) | 46.8% | 36.8% | 27.8% | 20.7% | 16.1% | 12.2% | 8.34% | 4.80% | 1.73% | 0.49% | 0.41% |
| | dual-source | 0.808 | 0.878 | 0.948 | 1.018 | 1.088 | 1.158 | 1.228 | 1.298 | 1.368 | 1.420 | 1.442 |
| high | flex-source | 0.583 | 0.653 | 0.723 | 0.793 | 0.863 | 0.933 | 1.003 | 1.073 | 1.143 | 1.213 | 1.275 |
| $c_f$ | (% ↓) | 27.8% | 25.6% | 23.7% | 22.1% | 20.7% | 19.4% | 18.3% | 17.3% | 16.4% | 14.6% | 11.6% |
| | base-source | 0.587 | 0.720 | 0.848 | 0.960 | 1.046 | 1.127 | 1.209 | 1.290 | 1.367 | 1.420 | 1.442 |
| | (% ↓) | 27.4% | 18.0% | 10.5% | 5.70% | 3.86% | 2.68% | 1.55% | 0.62% | 0.07% | 0.00% | 0.00% |

In Figure 2.4(a) and also the upper half of table 2.2, we compare the profit-to-go performance of dual-sourcing and single-sourcing when $c_b = 15$ and $c_f = 18$. We see that when the initial capacity position $\tilde{x}_n$ is low (below 20 in this case), dual-sourcing has a strictly better performance, particularly over the scenario of single-sourcing with the base supplier only. Specifically, when the initial capacity position is 0, which is the case when the manufacturer initiates the production, the expected total profit achieved under dual-sourcing is 87.9% higher than that achieved with single base sourcing. The performance gap closes as the initial capacity position increases and less additional capacity is needed. We also notice that when $c_f$ is close to $c_b$ as in this case, the profit difference between single flexible sourcing and dual-sourcing is within 3%. As we move to the other case when $c_f = 35$ (corresponding to Figure 2.4(b) and the lower half of Table 2.2), however, we see that dual-sourcing achieves considerably higher profit than both single base sourcing and single flexible sourcing (about 38% higher) when initial capacity position is low.

**Remark.** If the manufacturer initiates production with zero on-hand capacity, then having dual sources is always much more beneficial than having the single base source. A comprehensive analysis under widely varied cost parameters indicates that the advantage of having dual sources over having a single flexible source is also considerable if the flexible ordering cost is not too low and the initial capacity level is not too high.

### 2.5.3 The Value of Forecast Updates

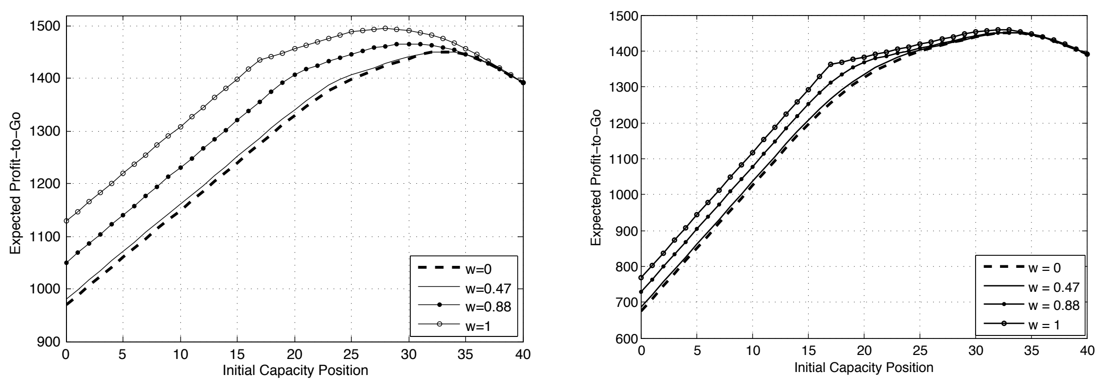
In this section we briefly explore the value of demand forecast updates under a dual-source capacity expansion setting. Intuitively, if we fix the distribution of period $n$ demand's total uncertainty, $\varepsilon_n^1 + \varepsilon_n^2$, then under a dual-sourcing strategy the expected profit should be higher when we increase the variance of $\varepsilon_n^1$, i.e., the portion of uncertainty that is captured by $\varepsilon_n^1$. If more demand information is contained in $\varepsilon_n^1$, then the flexible source will play a bigger role in terms of responding to the realization of the first market information instantaneously. There is some difficulty in designing the experiment. To facilitate programming, we want to use discrete uniform distribution with support on a consecutive integer range for both $\varepsilon_n^1$ and $\varepsilon_n^2$; however, it is hard to find pairs of distributions that have the above features while keeping the distribution of their summation unchanged. To simplify, let $U_1 \sim U[-4:1:4]$ with mean 0 and standard deviation 2.58, and $U_2 \sim U[-8:1:8]$ with mean 0 and standard deviation 4.90. Also, let $\tilde{U} := U_1 + U_2$, hence the distribution of $\tilde{U}$ can be denoted by Table 2.3 (with mean 0 and standard deviation 5.54):

Table 2.3: Distribution of $\tilde{U}$

| support | ±12 | ±11 | ±10 | ±9 | ±8 | ±7 | ±6 | ±5 | ±4 | ±3 | ±2 | ±1 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| probability | $\frac{1}{153}$ | $\frac{2}{153}$ | $\frac{3}{153}$ | $\frac{4}{153}$ | $\frac{5}{153}$ | $\frac{6}{153}$ | $\frac{7}{153}$ | $\frac{8}{153}$ | $\frac{9}{153}$ | $\frac{9}{153}$ | $\frac{9}{153}$ | $\frac{9}{153}$ | $\frac{9}{153}$ |

Now, define the *weight* of market information $\varepsilon_n^1$ to be $w := \frac{\sigma(\varepsilon_n^1)}{\sigma(\varepsilon_n^1 + \varepsilon_n^2)}$, where $\sigma(\cdot)$ is the standard deviation of the argument. We compare the profit performance of four different cases: $\varepsilon_n^1 = 0$, $\varepsilon_n^2 = \tilde{U}$ ($w = 0$); $\varepsilon_n^1 = U_1$, $\varepsilon_n^2 = U_2$ ($w \doteq 0.47$); $\varepsilon_n^1 = U_2$, $\varepsilon_n^2 = U_1$ ($w \doteq 0.88$); and finally, $\varepsilon_n^1 = \tilde{U}$, $\varepsilon_n^2 = 0$ ($w = 1$). Representative results are summarized in Figure 2.5 and Table 2.4.

As we increase the weight of $\varepsilon_n^1$ gradually from 0 to 1, we observe that for each state the expected profit also increases. More specifically, compared with the benchmark case where there is no demand forecast update ($w = 0$), the expected profits corresponding to $w = 0.47$, 0.88 and 1 are increased by approximately 1.6%, 7.9%, and 13.7%, respectively, for the high $c_f$ case (Figure 2.5(b)), and by 1.2%, 8.3%, and 16.4%, respectively, for the low $c_f$ case (Figure 2.5(a)), when the initial capacity and the modified backorder level are both 0. This nontrivial profit increase is attributed to the value of updated demand information.




(a) $c_b = 15,\ c_f = 18$ (b) $c_b = 15,\ c_f = 35$

Figure 2.5: Expected Profit under Different Weights of $\varepsilon_n^1$ ($n = 2,\ \tilde{y}_2 = 0$)

**Remark.** The flexible source can be used to capture both backorders from the previous period and to enable a response to the updated demand information for the current period. Therefore, the value of having dual-sourcing is largely dependent on the initial market information $\varepsilon_n^1$ that can be observed before the final demand is realized. The more information $\varepsilon_n^1$ contains, the higher profit dual-sourcing can achieve. To fully utilize the advantage of dual-source capacity expansion, the manufacturer should also improve the efficiency of its demand forecast updating process.

Table 2.4: The Expected Profit Under Different Weights of $\varepsilon_n^1$ ($\times 10^3$; $n = 2$; $\tilde{y}_2 = 0$; $\% \uparrow$ means percentage profit increase compared with the case of $w = 0$)

| initial capacity | | 0 | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $w = 0$ | 0.9696 | 1.0056 | 1.0416 | 1.0776 | 1.1136 | 1.1496 | 1.1856 | 1.2216 | 1.2576 | 1.2936 | 1.3296 |
| | $w = 0.47$ | 0.9801 | 1.0170 | 1.0530 | 1.0890 | 1.1250 | 1.1610 | 1.1970 | 1.2330 | 1.2690 | 1.3050 | 1.3410 |
| low | ($\% \uparrow$) | 1.17% | 1.13% | 1.09% | 1.05% | 1.02% | 0.99% | 0.96% | 0.93% | 0.90% | 0.88% | 0.85% |
| $c_f$ | $w = 0.88$ | 1.0503 | 1.0863 | 1.1223 | 1.1583 | 1.1943 | 1.2303 | 1.2663 | 1.3023 | 1.3383 | 1.3743 | 1.4069 |
| | ($\% \uparrow$) | 8.32% | 8.02% | 7.75% | 7.49% | 7.25% | 7.02% | 6.81% | 6.60% | 6.42% | 6.24% | 5.81% |
| | $w = 1$ | 1.1289 | 1.1649 | 1.2009 | 1.2369 | 1.2729 | 1.3089 | 1.3449 | 1.3809 | 1.4169 | 1.4419 | 1.4559 |
| | ($\% \uparrow$) | 16.4% | 15.8% | 15.3% | 14.8% | 14.3% | 13.9% | 13.4% | 13.0% | 12.7% | 11.5% | 9.50% |
| | $w = 0$ | 0.6751 | 0.7451 | 0.8151 | 0.8851 | 0.9551 | 1.0251 | 1.0951 | 1.1641 | 1.2271 | 1.2811 | 1.3258 |
| | $w = 0.47$ | 0.6861 | 0.7561 | 0.8261 | 0.8961 | 0.9661 | 1.0361 | 1.1061 | 1.1752 | 1.2376 | 1.2911 | 1.3349 |
| high | ($\% \uparrow$) | 1.63% | 1.48% | 1.35% | 1.24% | 1.15% | 1.07% | 1.00% | 0.95% | 0.86% | 0.78% | 0.69% |
| $c_f$ | $w = 0.88$ | 0.7282 | 0.7982 | 0.8682 | 0.9382 | 1.0082 | 1.0782 | 1.1482 | 1.2182 | 1.2837 | 1.3341 | 1.3683 |
| | ($\% \uparrow$) | 7.87% | 7.13% | 6.51% | 6.00% | 5.56% | 5.18% | 4.85% | 4.65% | 4.61% | 4.14% | 3.21% |
| | $w = 1$ | 0.7677 | 0.8377 | 0.9077 | 0.9777 | 1.0477 | 1.1177 | 1.1877 | 1.2577 | 1.3277 | 1.3697 | 1.3837 |
| | ($\% \uparrow$) | 13.7% | 12.4% | 11.4% | 10.5% | 9.70% | 9.03% | 8.46% | 8.04% | 8.20% | 6.92% | 4.37% |

## 2.6. Conclusion

In this chapter, we studied a dynamic dual-source capacity expansion problem with backorders, in which a manufacturer procures production capacity from two supply sources with consecutive leadtimes and different prices. After assuming an additive form of demand in terms of constant mean value and two random market information elements, we were able to derive a dynamic programming recursion with two state variables: the initial capacity position and the modified backorder level. Joint concavity holds for the objective function. However, unlike in dual-source inventory control problems or dual-source capacity expansion problems with lost sales, where a base-stock policy exists for the fastest two supply modes at optimality, we demonstrated that under dual-source capacity expansion with demand backlogging only the flexible (fast) orders follow a state-dependent base-stock policy. The base (slow) orders, on the other hand, follow a so-called partial-base-stock policy, where there exists a constant expand-to level only when the initial capacity position is below the base-stock level of the flexible source. We also obtained some monotonicity properties on the optimal (partial) base-stock levels. These results contribute to the literature on dynamic dual-source inventory and capacity management. In addition, we quantified the value of having dual sources and demand forecast updates during capacity planning through a brief numerical analysis.

We conclude that the dynamic dual-source capacity expansion problem with backorders is inherently different and more complex than its inventory counterpart as well as the dynamic dual-source capacity expansion problem with lost sales. Even in the simplest setting, i.e., under consecutive zero-one leadtimes, the optimal policy for this problem lacks structure. Given the challenges of capacity planning in capital-intensive industries and the underlying inefficiencies, it is thus critical to develop a better understanding of the complex balancing act of capacity procurement. Constructing solutions to handle the inherent tradeoffs will improve the capacity expansion process and will potentially save hundreds of millions of dollars in these industries.

# Chapter 3

# Dual-Mode Equipment Procurement Heuristic

## 3.1. Introduction

### 3.1.1 Motivation

Capacity planning is a complex balancing act, especially in the semiconductor industry. This industry is one of the most capital-intensive industries in the world; a single piece of semiconductor manufacturing equipment commonly costs tens of millions of dollars. Compounding the problem of high costs are the long leadtimes and the volatile consumer market. The order-to-production cycle for semiconductor manufacturing equipment can take up to 16 months, which exacerbates the difficulty of forecasting demand accurately. This chapter addresses the challenges of capacity planning in the semiconductor industry and describes our efforts at Intel to tackle these challenges by continuously improving the set of rules for Intel's engagement with the equipment suppliers.[1]

In the semiconductor industry, the marginal cost of unmet demand is considered to be significantly higher than the marginal cost of idle capacity (Fleckenstein 2004). Thus, despite the astounding costs, semiconductor firms often err on the side of having

[1] This chapter is a joint work with Feryal Erhun, Erik Hertzler, and Karl Kempf. A related paper is currently under revision at *Manufacturing & Service Operations Management*.

excess capacity and keep some equipment idle in order to not lose customer goodwill and loyalty. As an example, Figure 3.1 shows Intel's historical capacity purchases for six consecutive process technologies (denoted by $T_{-5}$ through $T_0$). The shaded areas in the figure display the difference between purchased capacity and realized demand at the peak of the demand curve. Consistently, all six generations suffered from excess capacity, the value of which is estimated to be several hundred million USD in capital depreciation per technology. Clearly, there are opportunities to improve this process.

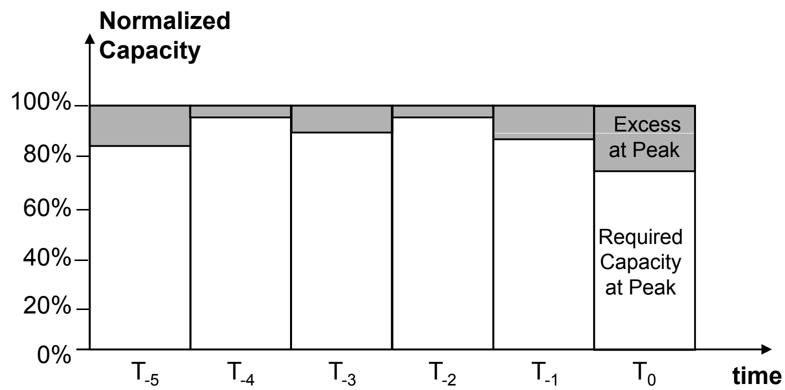



Figure 3.1: Purchased Capacity vs. Realized Demand at Intel

Traditionally, an easy way for firms with strong bargaining power to partially mitigate this capacity risk and avoid purchasing too much excess equipment has been through soft orders. To secure procurement contracts, suppliers have allowed firms to over-order capacity at the beginning of the planning horizon and to cancel some of the excess equipment without paying high penalties when more precise demand information becomes available, a process similar to the "phantom ordering" common in the personal computer and electronics industry (Lee *et al.* 1997, Cohen *et al.* 2003). As a result, equipment suppliers have carried most of the demand risk. However, this type of relationship is no longer sustainable. To keep pace with the technology requirements of maintaining Moore's Law (Moore 1965), the cost of capital equipment in the semiconductor industry has been steadily rising with no end in sight. Building a fab now costs in excess of 5 billion USD, up from 6 million USD in 1970 and around 2 billion USD in 2001 (Kanellos 2003). Hence, soft orders are costly to suppliers. In addition, the recent trend of supply consolidation (Armbrust 2009) has further increased suppliers' bargaining power and as a result, suppliers are becoming reluctant to carry demand risk through soft orders. On top of these facts, soft orders under most circumstances refer to orders that can be cancelled only BEFORE shipment. Even if allowed, they cannot totally eliminate over-capacity since there still can be huge uncertainty between capacity delivery and demand realization. Thus, the challenge for a semiconductor manufacturing firm is to derive innovative ways to order the right amount of equipment at the right time and price in advance of the demand realization to minimize unnecessary equipment purchases while maintaining a high level of service. Further, this goal should be achieved without pushing all of the risk onto suppliers.

The industry leader Intel and its suppliers are always working to refine the set of rules for their engagement. At Intel, the continuous improvement of capital acquisition and installation processes is an ongoing corporate priority. To be able to respond rapidly to changing customer demand for products while minimizing unnecessary capital expenditures is vitally important to both the semiconductor market and the firm itself. From Intel's perspective, demand uncertainties due to extremely long procurement leadtimes add too much idle capacity to the system and jeopardize the agility of the supply chain. Intel prefers to have tighter control of its capital supply chain by shortening equipment procurement leadtimes and improving the accuracy of the demand forecast used in capacity planning. In return for this flexibility, Intel is willing to take on some risk from its suppliers and is considering risk-sharing mechanisms. In this chapter, we propose a *dual-mode equipment procurement* (DMEP) model, which effectively addresses the above concerns by incorporating a fast supply mode (in addition to the regular procurement mode), a forecast revision mechanism, and a capacity reservation procedure. In designing DMEP, our goal is to guide Intel during the phases of equipment procurement. In particular, we are interested in the following questions: How can Intel quickly evaluate different flexible options during contract negotiations? Under what circumstances does the flexible mode create value for the firm? How much of the total capacity should be reserved through the flexible mode? When should this capacity be exercised?

The rest of the chapter is organized as follows: In Section 3.1.2, we briefly introduce the dual-mode equipment procurement model, which consists of three layers: a strategic *contract negotiation* layer, a tactical *reservation* layer, and an operational *execution* layer. Although motivated by Intel's continuous improvement efforts on equipment procurement, this model is versatile enough to be adapted by firms in other capital-intensive industries, such as the pharmaceutical and automotive industries. Section 3.2 provides a brief literature review for each layer of the model. Chapter 2 of this dissertation already shows that the execution layer of this model becomes intractable even under the simplest settings. Hence, we construct a heuristic approach for this problem in Section 3.3. In Section 3.4, we implement a detailed numerical analysis to quantify the value of the added flexibility that the dual-mode equipment procurement model provides. In particular, Section 3.4.3 looks at an extension that incorporates the firm's risk-averse attitude. Section 3.5 concludes the chapter and provides managerial insights. All proofs are presented in the appendix. Throughout the chapter, we use the terms *increasing* and *decreasing* in the weak sense; i.e., including equalities.

### 3.1.2 The Business Problem

As discussed in Section 3.1.1, Intel's goal is to continuously improve its equipment procurement strategies to support customer demand without over-purchasing costly capacity. Such a strategy should also embody fairness to its suppliers through a risk-sharing mechanism. To achieve this goal, we introduce the dual-mode equipment procurement (DMEP) model.

DMEP enables a firm to procure equipment from one supplier using two supply modes with complementary leadtimes and prices: a *base* mode (which is the regular procurement mode) that is less expensive but has a longer procurement leadtime $L_b$ and a *flexible* mode that is more expensive but has a shorter procurement leadtime $L_f$. The introduction of the flexible supply mode allows Intel to learn more about demand before committing to capacity. In return, to share the risk with its suppliers, Intel makes an up-front payment to secure certain base and flexible capacity levels ahead of time and faces prohibitive costs to cancel an order once it has been placed. As such, DMEP captures different phases of the relationship between Intel and its suppliers and is composed of three stages: contract negotiation, capacity reservation, and procurement execution (Figure 3.2).

During the *contract negotiation* stage, several years before the adoption of a new

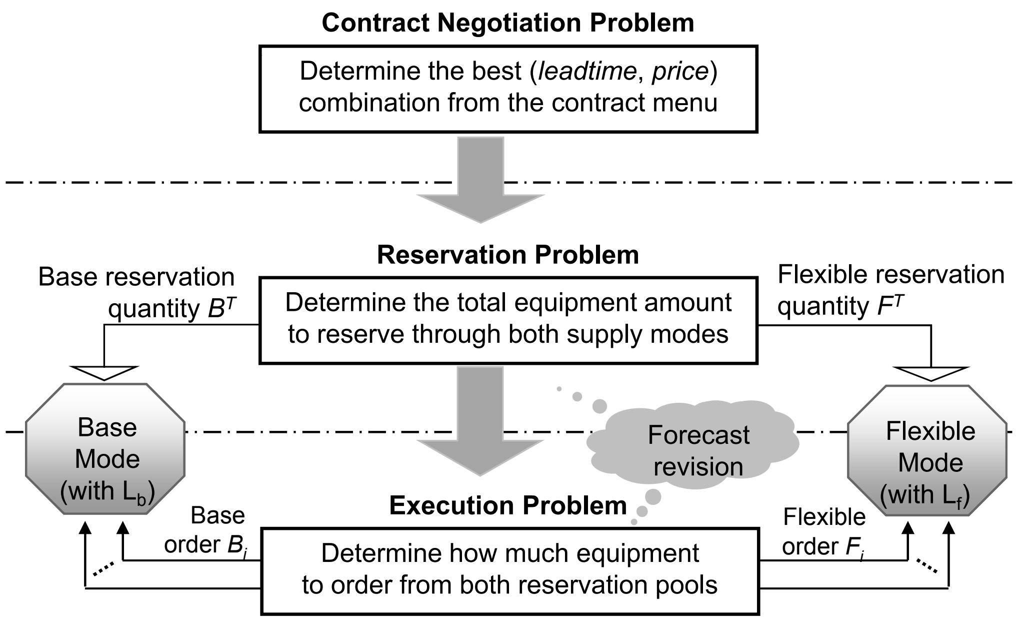



Figure 3.2: The Dual-Mode Equipment Procurement Model

process and the production of the necessary equipment, Intel and its supplier agree on the parameters of the supply modes. That is, they negotiate the leadtimes and prices (a reservation price and an execution price) of the base and flexible modes. During the *reservation* stage, several quarters before the planning horizon starts, Intel reserves the total equipment procurement quantities $B^T$ and $F^T$ from the base and flexible supply modes at unit reservation prices $r_b$ and $r_f$, respectively. These reservation quantities act as an upper bound on the *total* order volumes from the two supply modes in all subsequent periods. During the *execution* stage, in each period $n$, Intel orders specific amounts of equipment $B_{n+L_b}$ and $F_{n+L_f}$ (the subscript denotes the period when the ordered equipment arrives) from the two supply modes at unit execution prices $c_b$ and $c_f$, respectively, given the latest demand forecast as well as the realized demand information from the market. The execution algorithm follows a modified rolling-horizon process that will be explained in more detail later on.

With its three stages, DMEP evaluates various tradeoffs inherent in this setting. First, the contract negotiation stage balances accurate demand information with sourcing from expensive modes. Procuring from the flexible mode allows Intel to react to the market condition later with better demand information; however, the flexible mode is costlier than the base mode. Furthermore, Intel pays more for shorter procurement leadtimes, and choosing the right leadtime-cost combination is crucial. Secondly, the reservation stage balances flexibility (by reserving adequate capacity) with high reservation costs. By having a large reservation pool up front, Intel can guarantee the punctual delivery of its future orders; however, the reservation payment implies a high opportunity cost if the equipment is not ordered. Finally, the execution stage balances capacity holding costs with demand backordering costs. Early equipment procurement reduces the potential risk of having unsatisfied demand; however, it also leads to a higher capacity holding cost and a high opportunity cost if the equipment remains idle.

The goal of DMEP is to choose the appropriate leadtime-cost combinations, reservation levels, and execution quantities to maximize the expected total profit for Intel across the entire planning horizon. It is important that DMEP achieves its goals without pushing all of the risk onto the supplier. As such, the reservation payment enables the supplier to obtain capital to prepare its own production capacity and to help the supplier provide Intel with the guaranteed flexibility of ordering equipment when needed. That is, the reservation stage allows Intel to share risk with its supplier. Through the availability of a shorter leadtime mode, DMEP allows Intel to have tighter control of its capital supply chain. Through up front payments and commitments, DMEP is more fair to Intel's suppliers.

## 3.2. Literature Review

The contract negotiation stage of DMEP builds on the literature on supply contracts. We refer the reader to Cachon (2003). The contract that we study is closely related to the one analyzed in Yazlali and Erhun (2010). The authors investigate a dual-supply contract with minimum order quantity and maximum capacity restrictions. This contract provides supply chain partners an enhanced mechanism to share and manage demand uncertainty in the context of inventory management. We adopt a similar structure here for capacity procurement. The negotiation of price-leadtime combinations is also related to the literature on pricing and leadtime quotation. This literature often assumes that the buyer's order quantity is a function of the price and leadtime quoted by the supplier and thus focuses on the supplier's optimal price-leadtime decision (Palaka *et al.* 1998; Liu *et al.* 2007). We, on the other hand, take the buyer's perspective and provide the buyer a decision-support tool that it can use while negotiating the procurement price and leadtime with its supplier.

The reservation stage addresses a capacity expansion problem. The capacity expansion problem under single-sourcing has been extensively studied in the literature; see Van Mieghem (2003) for a review of this literature. Wu *et al.* (2005) provides a thorough review of the literature on capacity planning problems in the high-tech industry. DMEP is also closely related to the literature on procurement and options contracts (e.g., Bassok and Anupindi 1997; Bassok *et al.* 1999; Vaidyanathan *et al.* 2005). In particular, the reservation stage of DMEP builds on the paper by Vaidyanathan *et al.* (2005), which focuses on the capacity contracts at Intel and discusses capacity options to better enable the factory ramps. In this chapter, we extend and merge these two streams of research by studying capacity expansion and option contracts in a dual-mode setting.

The execution stage of DMEP is closely related to the dual-sourcing problem, except that here we assume the firm is sourcing from two service modes of the same supplier rather than from two different suppliers. The dual-sourcing problem has been studied extensively in the context of inventory since the early 1960s (Daniel 1963; Fukuda 1964; Whittemore and Saunders 1977; Scheller-Wolf and Tayur 1998; Feng *et al.* 2006; Yazlali and Erhun 2007, 2009). Dual-source inventory management has also been commonly adopted as an operational risk hedging strategy by firms in different industries for many years, e.g., Mattel (Johnson 2005) and HP (Billington and Johnson 2002). Despite the existence of literature on the dual-sourcing inventory problem, research concerning the dual-sourcing capacity procurement problem is scarce, with the exception of one recent paper (Chao *et al.* 2009); Chao et al. (2009) investigate a dual-source setting for the service industry where demand in excess of capacity is lost; they establish that the optimal capacity expansion policy with the fast source is base-stock. The authors further demonstrate that when the capacity obsolescence rate is deterministic, the optimal policy for capacity expansion through the slow source is also base-stock. Their model is restricted in the sense that: (1) it only deals with the lost-sale case, which is a simpler setting for capacity expansion problems; (2) it assume consecutive 0-1 leadtimes. The model we construct in this chapter tries to relax both of these assumptions.

Another relevant stream of literature is about rolling-horizon decision-making with information updates. Researchers in the 70's (Baker 1977, Baker and Peterson 1979) examined the effectiveness of rolling-horizon planning in the classic dynamic lot-sizing setting and concluded that rolling schedules are quite efficient. Some subsequent research (Bitran and Yanasse 1984, Bitran and Leong 1992) discusses deterministic approximations to stochastic production planning problems based on rolling-horizon concepts and quantifies the optimality gaps. Yildirim *et al.* (2005) extend the above stochastic production problem to a dual-source setting, where the manufacturer has both an in-house production facility and a subcontractor, with the same delivery leadtime but different costs and capacity limits. They demonstrate that the performance of a rolling-horizon deterministic equivalent model is very close to that of a benchmark threshold policy. Most of the literature in this area looks at an infinite period problem with a finite period rolling horizon; during each period, the information update only refers to the demand realization for the current period as well as the forecast for the additional period that has just entered the rolling window; the forecast information for all the overlapping periods are not updated. To incorporate a forecast updating mechanism (e.g., Graves *et al.* 1985, Eppen and Iyer 1997, Donohue 2000), Lian *et al.* (2010) investigate a rolling-horizon inventory replenishment model in which the buyer can update demand information and modify the previously committed order quantities; however, they limit their analytical discussion to a two-period model only. Compared to the extant literature, the modified rolling-horizon algorithm developed at the execution stage of our DMEP model has some unique features: First, it explicitly handles a dual-source procurement problem with nonconsecutive leadtimes and cross orders. Second, it is integrated with a systematical range forecast updating mechanism. Last, it is applied to a finite-period problem and the rolling window has a moving left boundary but a fixed right boundary.

In sum, by utilizing the concepts of capacity expansion, dual-sourcing, and rolling-horizon decision-making, we build a heuristic solution for DMEP, which will be described in detail in Section 3.3. We believe that in addition to its practical value, our work contributes to the OM literature and the research community in two ways:

1. We develop a comprehensive framework to structure the multi-stage decision hierarchy and capture the real-world dynamics and constraints for the equipment procurement problem.

2. We provide an efficient heuristic to solve the rather complex dual-source equipment capacity expansion problem with general leadtimes and demand backlogging, responding to the challenge raised in Chapter 2. The modified rolling-horizon algorithm can also be applied to solve other nonconsecutive-leadtime, multi-sourcing inventory control problems, which would otherwise be intractable.

## 3.3. Dual-Mode Equipment Procurement Heuristic

As discussed in Chapter 2 and Section 3.2, theoretically, the equipment procurement decision that the firm faces fits in the scope of a dual-source capacity expansion problem and could be formulated as a dynamic programming model. In practice, however, this may not be the best approach to take for two reasons. First, this is a complex problem and the dynamic programming model is rather difficult to solve. Even for the simplest setting where the leadtimes are consecutive with $L_b = 1$ and $L_f = 0$, Chapter 2 demonstrates that the dual-source capacity expansion problem with backorders does not have a well-behaved optimal policy. The general case with nonconsecutive leadtimes is even more complex. This complexity not only jeopardizes the possibility of finding any structural policy, but also sharply reduces the computational efficiency due to the famous curse of dimensionality. Second, although the dynamic programming model may help us identify the optimal equipment procurement strategy, it comes with strict assumptions, such as fixed and known distributions of all uncertain factors and consecutive leadtimes, which can hardly be justified in practice. Therefore, to avoid the above restrictions, we propose an open-loop simulation model as a heuristic approach. The goal of this approach is to provide Intel a fast and accurate decision-support tool with what-if capabilities that will guide the firm in answering the three questions we pose in Section 3.1.1.

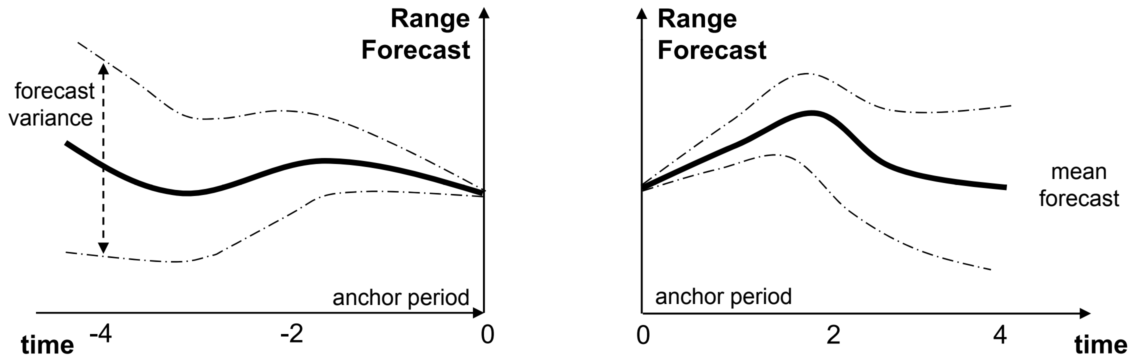
To map the structure presented in Figure 3.2, the DMEP heuristic algorithm consists of the same three layers: the outermost is the contract negotiation layer, which identifies the indifference curves of leadtime and price combinations so that the firm can pick the best alternative from the contract menu; the middle is the reservation layer, which calculates the optimal equipment reservation quantities $B^T$ and $F^T$ for the two supply modes; and the innermost is the execution layer, which determines the equipment order quantities $B_{n+L_b}$ and $F_{n+L_f}$ from the two supply modes in period $n$. Before elaborating on each of these three layers in more detail, we first consider the demand forecast revision process, which is one of the main drivers of the problem.

### 3.3.1 The Forecast Revision Mechanism

One of the major factors that complicate the equipment procurement decision is demand volatility. In a highly uncertain world, we need a detailed forecast revision model that can describe not only the variance associated with each period's demand, but also the process by which the firm continuously adjusts its anticipation of the future demand distribution based on the available information. Meanwhile, such a demand revision model would have the greatest value to the industry practice if it also observes a simple format that is convenient to implement.

Following the above guidelines, we modify the classical *martingale model of forecast evolution*, i.e., MMFE (Hausman 1969, Graves *et al.* 1986), by decomposing it into a mean evolution process and a variance evolution process. In particular, define $\mathcal{D}_n^m$ as the forecast made in period $m$ for the demand in period $n$ ($m = 1 - L_b, \cdots, N - L_f$; $n = 1, \cdots, N$; $m < n$); we use $\mu_n^m$ and $\sigma_n^m$ to represent the mean and the coefficient of variation (c.v.) associated with $\mathcal{D}_n^m$, and let $\mathcal{D}_n^m \sim \mathcal{N}(\mu_n^m, \sigma_n^m)^2$; $\mathcal{D}_n^n$ then refers to the realized demand of period $n$. We assume that the mean forecast evolution for a certain period $n$'s demand follows a markovian process with either an additive form: $\mu_n^m = \mu_n^{m-1} + \varepsilon_n^m$ ($m = 2 - L_b, \cdots, n-1$), where all the $\varepsilon_n^m$'s are independent random variables with zero mean, or a multiplicative form: $\mu_n^m = \mu_n^{m-1}\hat{\varepsilon}_n^m$ ($m = 2 - L_b, \cdots, n-1$), where all the $\hat{\varepsilon}_n^m$'s are independent positive random variables with mean value $E\hat{\varepsilon} = 1$. For both cases the initial mean forecast profile at period $1 - L_b$ is given by $\vec{\mu}^{1-L_b} = (\mu_1^{1-L_b}, \cdots, \mu_N^{1-L_b})$. The choice

[2] We modify the regular normal distribution notation by putting c.v. in the variance position.




(a) Forecast for anchor period's demand made at all previous periods

(b) Forecasts for all future periods' demand made at anchor period

Figure 3.3: Graphical Demonstration of the Forecast Revision Mechanism

between these two forms should depend on specific business environment and the company's empirical forecasting data. In a practical context, the multiplicative form is preferred due to two reasons (Hurley *et al.* 2007) despite its analytical complexity: first, the additive format may lead to negative demand values; second, industry forecasts are more frequently updated in a relative sense rather than an absolute sense. We further assume that the forecast variance evolution is predictable and the forecast c.v. is uniquely determined by the forecasting leadtime: $\sigma_n^m = f(n-m)$, where $f(\cdot)$ is an increasing function. We can choose different functional forms (e.g., linear, quadratic, logarithmic, etc.) for $f(\cdot)$ to reflect how the firm's forecasting ability changes with the forecasting leadtime.

Although this decomposed revision model may first seem like a notational complication to the standard MMFE method, it is actually easier to navigate in practice based on the feedback from Intel. Especially when compared to the original multiplicative MMFE process under which $\mathcal{D}_n = \mu_n^{1-L_b} \prod_{m=2-L_b}^{n} \hat{\varepsilon}_n^m$, our approach provides a technical simplification in the sense that we can now avoid calculating the cumulative distributions that involve the multiplication of several random variables. For firms that were used to point forecasting rather than range forecasting during the operational planning process, they can now update the mean forecast as if they were still doing point estimates, and then simply bundle it with the corresponding variance to generate a stochastic range forecast. We believe that our approach achieves a tradeoff between analytical rigorousness and managerial applicability. However, we want to point out that the overall DMEP framework is actually general enough to be combined with other forecast revision mechanisms as well.

Figure 3.3 graphically describes the demand forecast updating process, where period 0 is chosen as an anchor period and does not necessarily correspond to the starting period of the demand ramp. Figure 3.3(a) shows how forecasts for period-0 demand $\mu_0^m$ and $\sigma_0^m$ ($m \leq 0$) evolve as we approach period 0. The solid curve represents the mean forecast evolution path, and the interval between the dashed curves denotes the variance range, which shrinks as the forecasting leadtime decreases. That is, the forecast accuracy for a certain period's demand improves as one moves closer to that period in time. Figure 3.3(b) depicts the forecasts $\mu_n^0$ and $\sigma_n^0$ ($n \geq 0$) made in period 0 for all future periods' demand: the solid curve represents the mean forecast and the dashed interval denotes the variance range, which diverges as the forecasting leadtime increases. That is, the forecasting accuracy decreases as one forecasts further into the future.

Next, we explain the DMEP heuristic in greater detail. We first discuss the execution layer, then the reservation layer and the contract negotiation layer, since the former is a fundamental building block for the others.

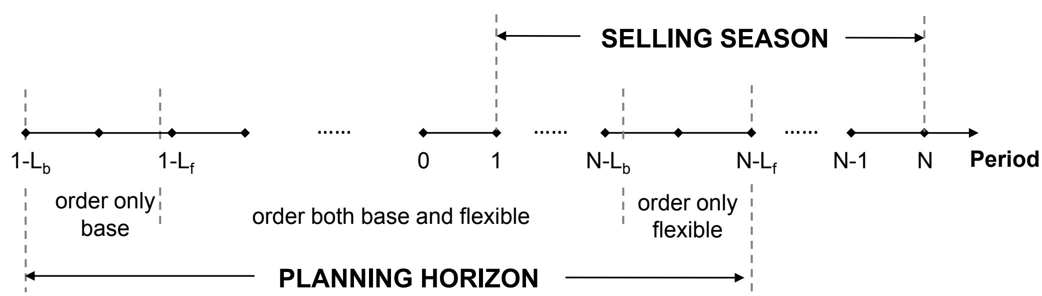



Figure 3.4: The Timeline of the Execution Problem

### 3.3.2 The Execution Problem

The execution module is the core of the DMEP heuristic. Given the reservation quantities $B^T$ and $F^T$, the execution module characterizes how the firm should place the base and the flexible orders in each period. At the beginning of each period, the firm obtains the latest demand realization and forecast updates for all future periods. Based on this information, all previously placed base and flexible orders, and the backorder quantity from the preceding period, the open-loop execution algorithm calculates the myopic optimal base and flexible order quantities for the remaining periods.

Figure 3.4 displays the timeline of the execution problem. The selling season of the product is $N$ periods and starts in period 1. To prepare for the demand ramp, the firm starts placing orders $L_b$ periods before the first demand realization and stops placing orders $L_f$ periods before the end of the selling season. Therefore, the length of the planning horizon is $N+L_b-L_f$ periods. The firm orders only from the base mode in periods $1-L_b, \cdots, -L_f$; it orders from both modes in periods $1-L_f, \cdots, N-L_b$; and it orders only from the flexible mode in periods $N-L_b+1, \cdots, N-L_f$. In periods $N-L_f+1$ to $N$, the firm simply satisfies the demand with the existing capacity.

Specifically, in period $m$ of the planning horizon ($m = 1-L_b, 2-L_b, \cdots, N-L_f$), the execution module solves the following stochastic optimization problem:

$$\underset{\vec{B}_{1,\cdots,N};\ \vec{F}_{1,\cdots,N}}{\text{maximize}} \quad \mathbb{E}_{d_{1\vee(m+1)},\cdots,d_N}\left[\sum_{i=1}^{N} \delta^i\{p_i s_i - c_b B_i - c_f F_i - c_h k_i\} - \delta^{N+1} c_u d_{N+1}^{rem}\right] \tag{3.3.1}$$

subject to: For each realized sample path $d_{1\vee(m+1)}, \cdots, d_N$

$$s_i = \min\{k_i, d_i + d_i^{rem}\} \quad \text{for } i = 1, \cdots, N \tag{3.3.2}$$

$$k_i \geq \psi d_i + d_i^{rem} \quad \text{for } i = 1 \vee (m+L_f), \cdots, N \tag{3.3.3}$$

$$k_i = k_{i-1} + B_i + F_i, \quad \text{for } i = 1, \cdots, N \text{ with } k_0 = 0 \tag{3.3.4}$$

$$d_i^{rem} = (d_{i-1} + d_{i-1}^{rem} - k_{i-1})^+, \quad \text{for } i = 1, \cdots, N \text{ with } d_1^{rem} = 0 \tag{3.3.5}$$

$$\sum_{i=1}^{N} B_i \leq B^T; \quad \sum_{i=1}^{N} F_i \leq F^T \tag{3.3.6}$$

$$\vec{B}_{1:N} \geq 0; \quad \vec{F}_{1:N} \geq 0 \tag{3.3.7}$$

$$\vec{B}_{1:m-1+L_b} = \vec{B}^*_{1:m-1+L_b}; \quad \vec{F}_{1:m-1+L_f} = \vec{F}^*_{1:m-1+L_f} \tag{3.3.8}$$

We define $x \vee y := \max(x, y)$ and use the notation $\vec{x}_{1:j}$ to represent the vector

$[x_1, x_2, \cdots, x_j]^T$. Additionally, $p_i$ is the profit margin for period $i$; $c_b$ the base mode execution price; $c_f$ the flexible mode execution price; $c_h$ the unit holding cost; $c_u$ the unit penalty cost for unmet demand after the planning horizon ends; $d_i$ the incoming demand in period $i$; $d_i^{rem}$ the unsatisfied demand remaining from period $i-1$; $s_i$ the actual sales quantity during period $i$; $k_i$ the cumulative capacity position at period $i$; $\psi$ the service level target; and $\delta$ the discount factor. $B_i^*$ and $F_i^*$ represent the optimal decisions that were already executed in previous periods. We note that the subscripts used for order quantities denote the period when the ordered equipment *arrives*.

The objective function maximizes the expected profit by considering the profit margin, equipment procurement costs, and inventory-related costs. The expectation is taken over all future demand and the execution cost is calculated when the orders arrive. We note that as our goal is to understand the strategic and tactical level capacity decisions, we suppress the operational level inventory problem and assume that production in a period will not exceed demand. Constraint (3.3.2) guarantees that sales in a given period cannot be larger than the demand or the supply; Constraint (3.3.3) forces the firm to satisfy at least $\psi$ percent of the incoming demand during each period after fulfilling the backorders. Constraints (3.3.4) and (3.3.5) enable state transitions for the capacity and backorders, respectively. Constraints (3.3.6) and (3.3.7) guarantee that the base and flexible orders are within their reservation limits and nonnegative. Finally, constraint (3.3.8) freezes the already executed orders before finding the optimal order quantities. One crucial feature of this rolling-horizon algorithm is that, in period $m$, only the orders $B_{m+L_b}^*$ and $F_{m+L_f}^*$ (if $m + L_f \geq 1$) will actually be placed, and the rest of the decisions will be postponed to future periods.

To reflect the industry practice that we are investigating, we make several modeling assumptions. First, we analyze a setting where the unmet demand during a certain period is backordered (instead of lost, which is assumed in most of the capacity planning literature). Instead of defining a unit penalty cost per period, we penalize the backordered demand in two ways: ($i$) at the end of the planning horizon, any remaining unsatisfied demand incurs a terminal penalty; and ($ii$) the profit margin of the product is decreasing over time, which implies that there is a loss of revenue associated with backordering. Therefore, satisfying the demand earlier is always preferable if it is possible. Second, the on-hand equipment capacity incurs a unit holding cost, which may include the opportunity cost of investment or costs such as a utility fee, maintenance expenditure, or cost of floor space. Third, investment in the equipment capacity is irreversible; i.e., capacity contraction is not allowed. Fourth, the firm is risk-neutral and maximizes its expected profit. Finally, the raw material inventory is always sufficient for the production process, and we only concentrate on the equipment procurement decisions.

| planning horizon | selling season: 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| -3 | $\mu_1^{-3}\sigma_4$ | $\mu_2^{-3}\sigma_5$ | $\mu_3^{-3}\sigma_6$ | $\mu_4^{-3}\sigma_7$ | $\mu_5^{-3}\sigma_8$ | $\mu_6^{-3}\sigma_9$ |
| | **B1*** | B2 | B3 | B4 | B5 | B6 |
| | F1 | F2 | F3 | F4 | F5 | F6 |
| -2 | $\mu_1^{-2}\sigma_3$ | $\mu_2^{-2}\sigma_4$ | $\mu_3^{-2}\sigma_6$ | $\mu_4^{-2}\sigma_6$ | $\mu_5^{-2}\sigma_7$ | $\mu_6^{-2}\sigma_8$ |
| | B1 | **B2*** | B3 | B4 | B5 | B6 |
| | F1 | F2 | F3 | F4 | F5 | F6 |
| -1 | $\mu_1^{-1}\sigma_2$ | $\mu_2^{-1}\sigma_3$ | $\mu_3^{-1}\sigma_4$ | $\mu_4^{-1}\sigma_5$ | $\mu_5^{-1}\sigma_6$ | $\mu_6^{-1}\sigma_7$ |
| | B1 | B2 | **B3*** | B4 | B5 | B6 |
| | **F1*** | F2 | F3 | F4 | F5 | F6 |
| 0 | $\mu_1^{0}\sigma_1$ | $\mu_2^{0}\sigma_2$ | $\mu_3^{0}\sigma_3$ | $\mu_4^{0}\sigma_4$ | $\mu_5^{0}\sigma_5$ | $\mu_6^{0}\sigma_6$ |
| | B1 | B2 | B3 | **B4*** | B5 | B6 |
| | F1 | **F2*** | F3 | F4 | F5 | F6 |
| 1 | ***Dmd1*** | $\mu_2^{1}\sigma_1$ | $\mu_3^{1}\sigma_2$ | $\mu_4^{1}\sigma_3$ | $\mu_5^{1}\sigma_4$ | $\mu_6^{1}\sigma_5$ |
| | B1 | B2 | B3 | B4 | **B5*** | B6 |
| | F1 | F2 | **F3*** | F4 | F5 | F6 |
| ... | | | | | | |

Figure 3.5: A Tabular Demonstration of the Execution Heuristic. $L_b = 4$, $L_f = 2$; $\sigma_l$ denotes the forecast variance corresponding to a forecasting leadtime of $l$ periods; $*$ current period decision; _ previous decisions; grey decisions not to be executed; and *Dmd* realized demand

To illustrate the firm's decision-making process more clearly, we present the execution heuristic algorithm using a concise tabular format in Figure 3.5. We assume that there are 6 periods in the selling season, the base mode leadtime is 4 periods, and the flexible mode leadtime is 2 periods. The horizontal axis in the table represents the entire selling season from period 1 to period 6. Each horizontal group below the top line contains the demand information and decision profiles corresponding to a decision period, which is labeled on the vertical planning horizon axis.

At the beginning of the planning horizon, period −3, the firm obtains the latest demand forecast information ($\mu$'s and $\sigma$'s) for period 1 to period 6. The algorithm runs the aforementioned stochastic optimization to maximize the expected total profit across the entire planning horizon under the current demand information, taking the six base orders ($B_1$ to $B_6$) and six flexible orders ($F_1$ to $F_6$) as the decision variables. Once the optimal order quantities are obtained, only the base order $B_1$ for period 1 is placed since, given the leadtimes of the two supply modes ($L_b = 4,\ L_f = 2$), the firm only needs to commit to $B_1$ at period −3; the rest of the decisions ($B_2$ to $B_6$, $F_1$ to $F_6$) are postponed until later. At the beginning of period −2, the firm obtains the updated forecast information (new $\mu$'s and $\sigma$'s); it then solves a new stochastic optimization to maximize the horizon-wide expected profit under the newly obtained demand information, fixing $B_1$ and taking $B_2$ to $B_6$ and $F_1$ to $F_6$ as the decision variables. Similarly, at period −2 only the decision $B_2$ needs to be executed and the rest of the decisions are left for later periods. Following this logic, $B_3$ and $F_1$ will be executed in period −1, etc. One can imagine that if $\mu_1^{-1}$ is much higher than $\mu_1^{-2}$ (demand forecast shock), then a positive $F_1$ will be committed at period −1. This rolling-horizon decision-making process continues until period 4, when the final order $F_6$ is committed. It is important to emphasize that starting from period 1, the actual realized demand is treated as part of the updated demand information, and should be taken into consideration when calculating the total expected profit.

We solve the above stochastic program using the standard sample average approximation method based on Monte Carlo simulation (Shapiro 2008). We are able to consider a large number of samples, which improves the accuracy of the solution, because the problem is easy to execute:

**Proposition 3.3.1.** *Assuming that demand $d_n$ is discrete and takes finitely many values, the above stochastic programming model can be converted into a linear programming model.*

The execution algorithm provides a handy roadmap that indicates when and how much to order from the two supply modes. It captures the evolution of demand information and enables last-minute decision-making. As opposed to the restricted focus of some rule-of-thumb approaches such as executing flexible orders only during the peak period, the potential use of the flexible mode is evaluated in each period of the planning horizon. On the flip side, the execution algorithm is myopic in the sense that it finds the "optimal" ordering scheme based on the current best information only, without considering the opportunity to make contingent decisions based on the actual realized demand at each stage. Fortunately, this disadvantage of myopia is greatly mitigated by the rolling-horizon nature of the algorithm.

**Proposition 3.3.2.** *When $c_f > c_b$, and $B^T$ and $F^T$ are not binding, the following are true: (i) flexible orders are only likely to be placed for periods for which a base order has already been committed; (ii) for stage m's problem, the number of free*[3] *base decision variables is $\Xi_b = [N+1-(m+L_b)]^+$, and the number of free flexible decision variables is $\Xi_f = (L_b - L_f) \wedge (m + L_b - 1) \wedge [N + 1 - (m + L_f)]^+$.*

Proposition 3.3.2 above says that when the flexible execution price is higher than the base execution price, which is indeed the practical situation under most business contexts, the execution problem can be further simplified through a reduction in the number of decision variables associated with the flexible mode. This again increases the computational efficiency of the execution problem.

### 3.3.3 The Reservation Problem

As equipment suppliers gain more bargaining power due to the trend of supply consolidation, simply letting the supplier bear the major procurement risk is no longer sustainable. The reservation module of the DMEP heuristic, therefore, functions as a mechanism for risk-sharing between the firm and its equipment supplier. The firm, by paying an up-front reservation fee, shares the risk of capacity building and installation with the supplier and enjoys the guaranteed delivery of equipment in return.

Determining how much capacity to reserve from the two supply modes, especially the flexible mode, is based on a tradeoff between the reservation cost and the potential benefits from the guaranteed flexibility. Specifically, the optimal reservation quantities $B^T$ and $F^T$ are determined according to a scenario analysis of the potential future demand profiles. Intuitively, if the future demand scenario involves no uncertainty, then flexibility has no value; it is never optimal to order from the expensive flexible mode. In contrast, if the future demand scenario is highly uncertain and the

[3] "Free" here means that the decision variable is not subject to an equality constraint.

demand mean forecast is very likely to be modified during the updating process, then flexibility has a high value and we should expect a higher flexible reservation level. In general, the reservation quantities maximize the expected horizon-wide profit over all possible demand scenarios.

More precisely, the reservation algorithm determines the optimal $B^T$ and $F^T$ based on a Monte Carlo simulation performed on the mean forecast evolution trajectories. It generates mean forecast evolution paths according to the forecast revision mechanism introduced in Section 3.3.1. Assuming that the mean forecast for the demand in different periods evolves according to a Markovian process, we then have

$$\begin{aligned} & P(\mu_1^m, \cdots, \mu_N^m | \mu_1^{m-1}, \mu_1^{m-2}, \cdots, \mu_N^{m-1}, \mu_N^{m-2}, \cdots) \\ = \;& P(\mu_1^m, \cdots, \mu_N^m | \mu_1^{m-1}, \cdots, \mu_N^{m-1}), \quad \text{for} \quad m < 1 \end{aligned} \tag{3.3.9}$$

where $P(\cdot)$ is the probability mass function if $\mu$ takes discrete values and the probability density function if $\mu$ takes continuous values. Therefore, at the beginning of the planning horizon, given the initial demand forecast profile $\vec{\mu}^{1-L_b}$ for the entire horizon and $P(\cdot)$, the algorithm enumerates a large number of possible mean forecast evolution paths. For each path it calls the execution module to calculate the specific order quantities as well as the expected horizon-wide profit. The algorithm then chooses the reservation quantities $B^T$ and $F^T$ that maximize the average total profit across the entire planning horizon.

Equation (3.3.10) formulates the reservation problem mathematically. We first denote the optimal value function of the period-$m$ execution problem (3.3.1)-(3.3.8) as $J^m(B^T, F^T, \vec{\mu}^m, \vec{\sigma}^m)$ by decomposing the demand information $d$ into its two components $\mu$ and $\sigma$. We choose $B^T$ and $F^T$ to maximize the expected horizon-wide profit $J^{N-L_f}$ since period $N - L_f$ is the last period during which a decision can be made. The stochastic optimization is given as

$$\max_{B^T \geq 0; F^T \geq 0} \quad \mathbb{E}_M \left[ J^{N-L_f}(B^T, F^T, M, \Sigma | \vec{\mu}^{1-L_b}) \right] - r_b B^T - r_f F^T \tag{3.3.10}$$

where the expectation is taken with respect to the mean forecast evolution space $M$:

$$M = \{\vec{\mu}^m : \vec{\mu}^m = (\mu_{1\vee(m+1)}^m, \cdots, \mu_N^m), \; m = 1 - L_b, \cdots, N - L_f\}. \tag{3.3.11}$$

As we expressed in equation (3.3.9), the mean forecast profile $\vec{\mu}^m$ follows a Markov process with initial state $\vec{\mu}^{1-L_b}$ and transition probability space $P(\cdot)$, such that a particular realization of $M$ occurs with probability

$$P(M|\vec{\mu}^{1-L_b}) = P(\vec{\mu}^{2-L_b}|\vec{\mu}^{1-L_b})P(\vec{\mu}^{3-L_b}|\vec{\mu}^{2-L_b})\cdots P(\vec{\mu}^{N-L_f}|\vec{\mu}^{N-1-L_f}). \quad (3.3.12)$$

Finally, the variance forecast evolution space is

$$\Sigma \;=\; \left\{\vec{\sigma}^m : \vec{\sigma}^m = (\sigma^m_{1\vee(m+1)}, \cdots, \sigma^m_N),\ m = 1 - L_b, \cdots, N - L_f\right\}; \quad (3.3.13)$$

$\Sigma$ is actually deterministic based on our assumption that the forecasting variance is uniquely decided by the forecasting leadtime. Proposition 3.3.3 below establishes the concavity and coerciveness of the reservation problem and thus the existence of finite optimal reservation quantities for the two supply modes. Hence, the reservation problem can be solved using either an optimization software or a search algorithm.

**Proposition 3.3.3.** *The objective function in (3.3.10) is concave and coercive in* $(B^T, F^T)$.

The reservation algorithm helps the firm determine the optimal amount of equipment to reserve for both supply modes. It builds on the philosophy of scenario analysis and chooses the reservation quantities that guarantee the maximum expected return to the firm. By adjusting the mean forecast transition probability $P(\cdot)$, we can easily create different demand scenarios; hence the reservation algorithm is applicable to a wide range of business settings.

**Proposition 3.3.4.** *The total reservation quantity* $B^T + F^T$ *increases in the service level parameter* $\psi$*, the mean evolution jump size* $\beta$*, and the demand variance coefficient* $\gamma$*.* [4]

The above proposition says that the firm tends to make more aggregate reservations when the service level requirement goes up or when the market becomes more volatile. This makes intuitive sense since for the semiconductor industry the cost of underage is higher than the cost of overage. In Section 3.4.2 we will discuss how the

[4] The formal definition of $\beta$ and $\gamma$ will be given in Section 3.4.1.

individual $B^T$ and $F^T$, as well as the ratio $\frac{F^T}{B^T+F^T}$, respond to the change of these parameters.

### 3.3.4 The Contract Negotiation Problem

During the contract negotiation stage, facing a contract menu potentially suggested by the equipment supplier, the firm determines the leadtimes $(L_b, L_f)$ of the two delivery modes as well as the unit reservation $(r_b, r_f)$ and execution $(c_b, c_f)$ prices associated with these leadtimes. Our algorithm involves a simple sensitivity analysis: for different (*leadtime, price*) combinations, we run the reservation and execution heuristic and obtain the corresponding expected horizon-wide profits. The decision-maker can then choose the (*leadtime, price*) pair from the contract menu that leads to the highest expected return for the firm.

The strength of this part of the heuristic is that it helps the firm make the optimal strategic level decision by considering potential tactical and operational level contingencies. Thus, the three stages of DMEP constitute a stable decision-support pyramid, in which the decisions are made in a top-down sequence while the underlying algorithm follows an embedded bottom-up order.

## 3.4. DMEP as a Decision-Support Tool

In this section, we revisit the questions that we asked in Section 3.1.1 and illustrate our approach to them with numerical examples. We first provide the parameter values that we use for the numerical examples. We then explore the value of DMEP as a decision-support tool with an emphasis on these three questions. We implemented the DMEP heuristic using the convex optimization tool CVX (http://cvxr.com/cvx/) in Matlab.

### 3.4.1 Parameter Values for the Numerical Examples

The parameter values we use in our examples are based upon the semiconductor business environment. The values presented here are either publicly available or have been normalized to protect the firm. For all numerical samples, the selling season is

$N = 6$ quarters. The profit margin (including the equipment cost) of a single chip is $p_0 = \$33.75$ before the first quarter of the life cycle starts; the profit margin $p_t$ in quarter $t$ satisfies an exponential decreasing formula: $p_t = p_0 e^{-\alpha t}$ (Leachman 2007) where the coefficient $\alpha$ is equal to 0.23, roughly implying that the margin decreases by 50% per year. The unit penalty cost for unmet demand at the end of the selling season is $C_p = \$50$. The service level constraint is 95%. Each piece of equipment can process 12,000 wafers per quarter, and each wafer can be further sawed into 1,425 chips. The discount factor per quarter is $\delta = 0.96$.

In this setting, the leadtime and price for the base mode are usually fixed. However, suppliers offer a menu of contracts for the flexible mode, which commonly includes the available flexible leadtime options as well as the price associated with each leadtime. The shorter the flexible leadtime, the higher the flexible price. Using this fact, we set the base leadtime $L_b$ to 4 quarters and the total equipment price associated with the base mode $p_e^b$ to \$25 million. We impose a minimum base reservation price of $r_b = \$0.1$ million to eliminate the trivial case of infinite base reservation quantity. The base execution price is thus given as $c_b = p_e^b - r_b$. We choose equipment with a long leadtime as these equipments are the bottleneck in capacity planning and are usually the ones which are very expensive. Hence they are difficult to manage and therefore the target product of DMEP. The price associated with the flexible mode is determined by two parameters: the equipment price increase ratio $\theta$ and the flexible reservation price ratio $\lambda$. Namely, the total equipment price of the flexible mode is $p_e^f = \theta p_e^b$ where $\theta \geq 1$ since a faster mode implies a shorter preparation leadtime for the supplier and hence a higher supply cost. The reservation price of the flexible mode is $r_f = \lambda p_e^f$. The remaining $(1 - \lambda)$ portion together with a fast shipment premium of \$50,000 is paid as the flexible execution price; that is, $c_f = (1 - \lambda)p_e^f + \$50,000$.

For the forecast revision process, we make some additional assumptions. First, the mean forecast evolution follows a multiplicative form ($\mu_n^m = \mu_n^{m-1} \varepsilon_n^m$), which can better reflect the fact that forecast error is usually proportional to the forecasted value. In particular, we introduce a mean-adjustment factor $\beta$ that can be tailored to each firm's own forecasting history, and assume that $\varepsilon_n^m = 1 + \beta, 1, 1 - \beta$ with equal probability $\frac{1}{3}$. The mean forecasts for demand in different periods evolve independently. Second, we assume the coefficient of variation $\sigma_n^m$ is linearly increasing in the forecast leadtime; i.e., $\sigma_n^m = \gamma \times 0.01(n-m)$. We note that $\beta$ controls the solid path in Figure 3.3 and $\gamma$ controls the width of the dashed variance interval. Last, demand in each period is truncated normal.

Table 3.1 displays the initial demand forecast that is provided by Intel. The initial forecast is symmetric with periods 3 and 4 being peaks. Using this initial forecast, we study three different forecast scenarios. In the first scenario, the initial forecast is not adjusted and the realized demand in each period matches the corresponding mean forecast. The only randomness in this system is the variance coefficient $\gamma$. Since the mean forecast evolution process is degenerative and hence $\beta = 0$ for all periods in this case, we call this scenario the *stationary demand* scenario. The second scenario investigates demand *forecast shocks*: during period $-1$ (1), the forecast for the mean demand in period 1 (3) is adjusted upwards. The third scenario corresponds to a situation with demand *realization shock*: the initial mean forecast profile $\vec{\mu}_{1,\cdots,N}^{-3}$ is not updated in periods $-2$, $-1$ and 0; in period 1, however, the actual realized demand is 14,000, which is much higher than the previous mean forecast.

Table 3.1: Nonstationary Demand Forecast Scenarios (unit: wafer-start-per-week)

| Scenarios | | period 1 | period 2 | period 3 | period 4 | period 5 | period 6 |
|---|---|---|---|---|---|---|---|
| Initial forecast | $\vec{\mu}_{1,\cdots,6}^{-3}$ | 4,788 | 9,577 | 14,365 | 14,365 | 9,577 | 4,788 |
| Stationary demand | $\vec{\mu}_{1,\cdots,6}^{-3}$ | 4,788 | 9,577 | 14,365 | 14,365 | 9,577 | 4,788 |
| Forecast shock | $\vec{\mu}_{1,\cdots,6}^{-1}$ | **7,000** | 9,577 | 14,365 | 14,365 | 9,577 | 4,788 |
| | $\vec{\mu}_{1,\cdots,6}^{1}$ | 6,850 | 9,577 | **19,320** | 14,365 | 9,577 | 4,788 |
| Realization shock | $\vec{\mu}_{1,\cdots,6}^{1}$ | **14,000** | 9,577 | 14,365 | 14,365 | 9,577 | 4,788 |

### 3.4.2 The Value of DMEP as a Decision-Support Tool

**Under what circumstances does the flexible mode create value for the firm?**

To capture the settings under which the flexible mode will create value for Intel, we first investigate the optimal equipment procurement decisions given the reservation quantities $B^T$ and $F^T$. Figure 3.6 illustrates the settings under which the flexible mode is utilized as a risk-hedging channel by summarizing our findings for the scenarios discussed in Table 3.1. The horizontal axis denotes the ratio between the

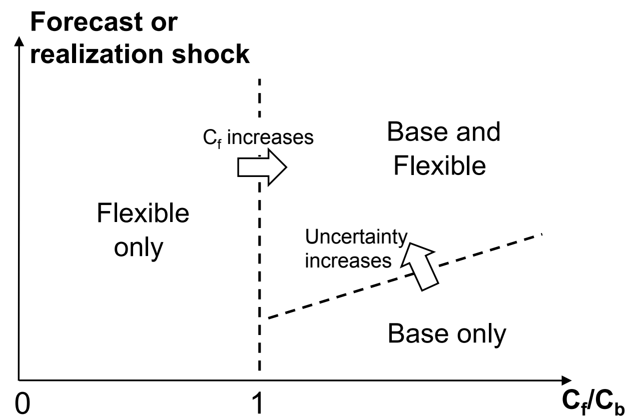
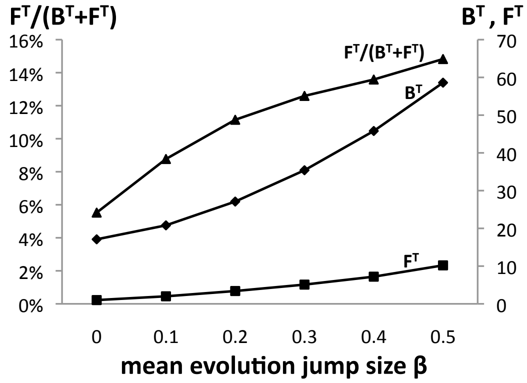


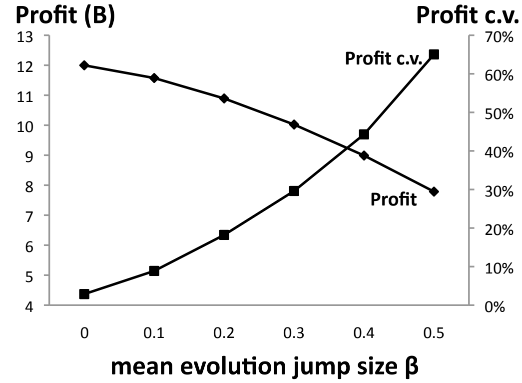

Figure 3.6: Impacts of Price and Demand Property on the Ordering Policy (illustration)

flexible execution price $c_f$ and the base execution price $c_b$; the vertical axis denotes the size of the forecast shocks and/or the size of the demand realization shocks. The entire plane can be divided into three different regions based on the value of the two coordinates: If $c_f/c_b \leq 1$, then the firm procures equipment only from the flexible mode, since it is not only faster but also cheaper. If $c_f/c_b > 1$ and the forecast and realization shocks are small enough, then it is optimal to procure only from the base mode. That is, in the stationary demand scenario or when shocks are small, cost is the only critical parameter: the firm prefers to single source using the less expensive mode as long as there is available reservation quantity. When there is forecast shock (previous demand forecast being adjusted upwards) or realization shock (extremely large demand realization), however, this threshold policy ceases to apply. If $c_f/c_b > 1$ and the forecast or realization shocks surpass a certain threshold value, then both the base mode and the flexible mode are used. Now the flexible mode adds value with its shorter leadtime even when it is more expensive. The firm can wait until the last minute to learn more about demand before placing orders and thus maintain an agile environment.

At the tactical capacity reservation level, Figure 3.7 and 3.8 demonstrate that as the demand risk increases (i.e., either the mean evolution jump size $\beta$ or demand variance coefficient $\gamma$ increases), the firm relies on the flexible mode more heavily. The flexible mode is especially valuable in dealing with demand realization shocks since

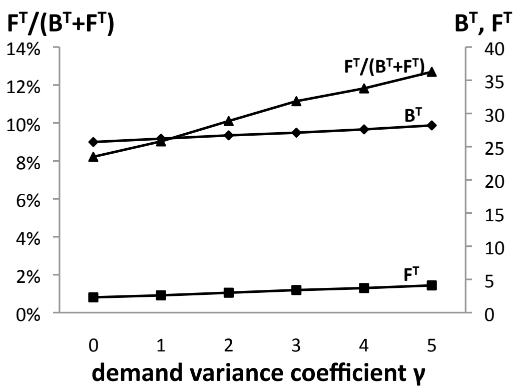



(a) Impact of $\beta$ on the reservation levels

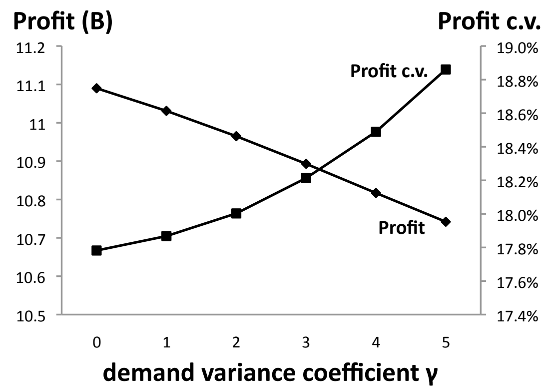



(b) Impact of $\beta$ on profit and its variability

Figure 3.7: Impact of Mean Evolution Jump Size $\beta$ ($\theta = 1.3,\ \lambda = 0.15,\ \gamma = 3,\ \psi = 0.95,\ L_f = 2$)



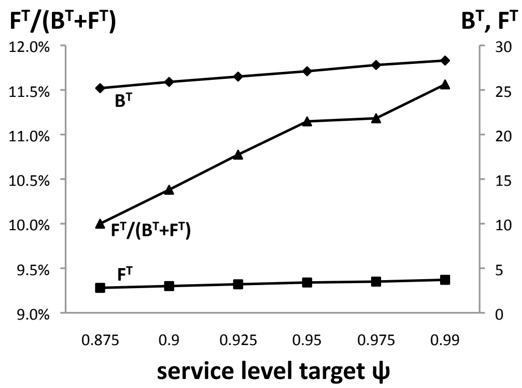

(a) Impact of $\gamma$ on the reservation levels



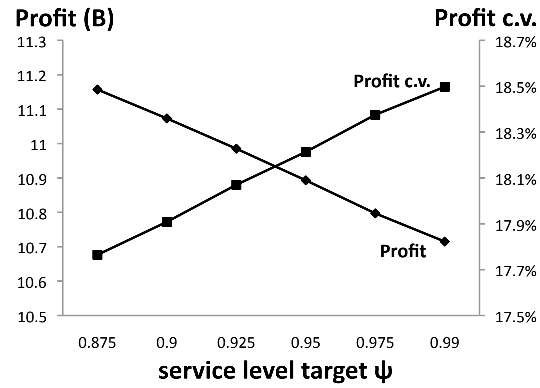

(b) Impact of $\gamma$ on profit and its variability

Figure 3.8: Impact of Demand Variance Coefficient $\gamma$ ($\theta = 1.3,\ \lambda = 0.15,\ \beta = 0.2,\ \psi = 0.95,\ L_f = 2$)

the firm's contingencies are very limited in that case. Finally, Figure 3.9 demonstrates that as we gradually increase the service level constraint from 87.5% to 99%, the value of the flexible mode increases. Note that the expected profit associated with a 99% service level is almost 200 million dollars less than that associated with a 95% service level. This is by no means a small compromise, and firms should consider this tradeoff when making service level decisions.

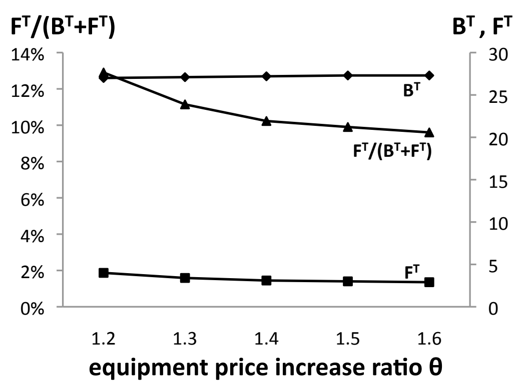



(a) Impact of $\psi$ on the reservation levels

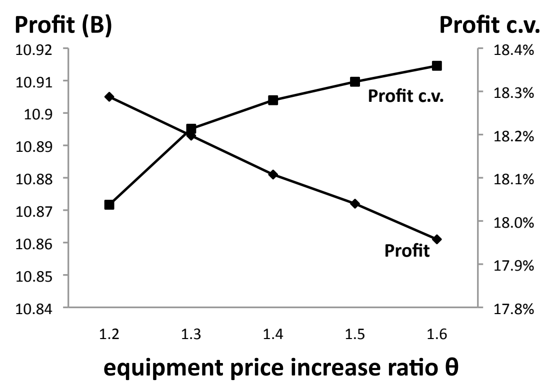



(b) Impact of $\psi$ on profit and its variability

Figure 3.9: Impact of Service Level Target $\psi$ ($\theta = 1.3,\ \lambda = 0.15,\ \beta = 0.2,\ \gamma = 3,\ \psi = 0.95,\ L_f = 2$)




(a) Impact of $\theta$ on the reservation levels

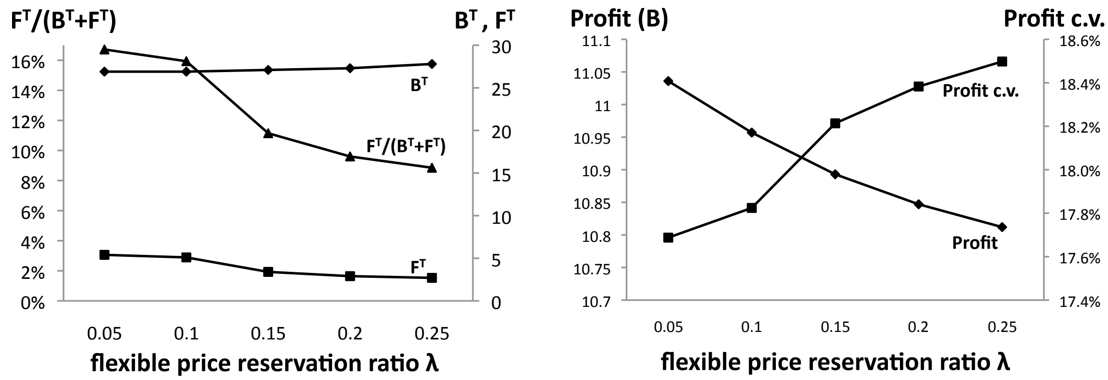



(b) Impact of $\theta$ on profit and its variability

Figure 3.10: Impact of Equipment Price Increase Ratio $\theta$ ($\lambda = 0.15,\ \beta = 0.2,\ \gamma = 3,\ \psi = 0.95,\ L_f = 2$)

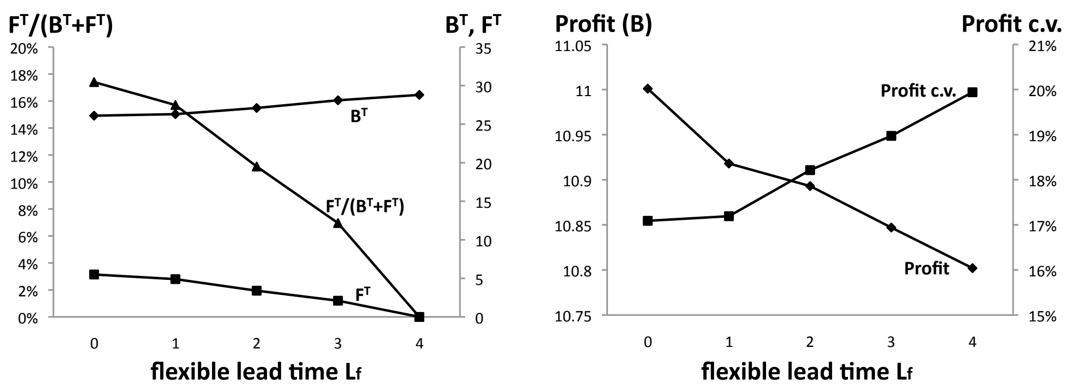
**How much of the total capacity should be reserved through the flexible mode? When should this capacity be exercised?**

We note that the conditions for dual-mode procurement correspond to the business environment in which Intel operates; i.e., the forecast and realization shocks, the forecast uncertainty, and service levels are all high. Thus, dual-mode procurement




(a) Impact of $\lambda$ on the reservation levels (b) Impact of $\lambda$ on profit and its variability

Figure 3.11: Impact of Flexible Price Reservation Ratio $\lambda$ ($\theta = 1.3,\ \beta = 0.2,\ \gamma = 3,\ \psi = 0.95,\ L_f = 2$)




(a) Impact of $L_f$ on the reservation levels (b) Impact of $L_f$ on profit and its variability

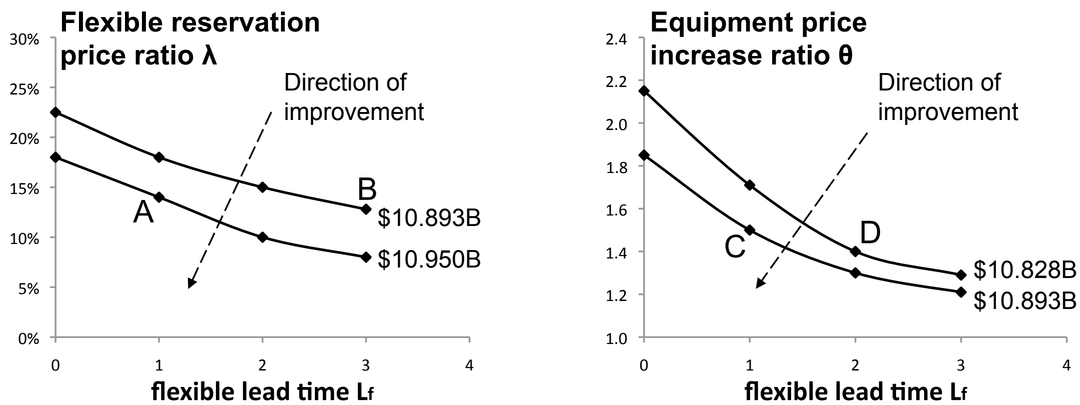
Figure 3.12: Impact of Flexible Mode Leadtime $L_f$ ($\theta = 1.3,\ \Lambda = 0.15,\ \beta = 0.2,\ \gamma = 3,\ \psi = 0.95$)

should be seriously considered in this industry, which has traditionally used single-mode procurement. The notion that flexibility is only necessary for peak periods is also a misconception. Actually, the flexible mode may be optimally deployed whenever there is a forecast shock or a realization shock, both of which occur frequently during the ramp-up stage of the product life cycle, instead of just at the peak. Although the value of the flexible mode decreases as its total price (i.e., $\theta$) and/or reservation price (i.e., $\lambda$) increase, as long as the flexible leadtime ($L_f = 2$) is significantly shorter than the base one, the firm continues to reserve more than 8% of its total capacity through the flexible mode even when the flexible mode is 60% more expensive (Figure 3.10) or when the firm has to pay 25% up front (Figure 3.11). When the flexible leadtime increases, however, the value of flexibility dramatically decreases and the firm tends to depend more on the base mode (Figure 3.12).

For all the above examples we also record the corresponding profit variability in terms of the coefficient of variation. It turns out as cost, uncertainty, service level, or flexible leadtime increases, the expected total profit decreases while the profit variability increases.

**How can Intel quickly evaluate different flexible options during contract negotiations?**

One efficient way for Intel to select offers from the contract menu is to compare the position of different leadtime and price combinations for the flexible mode on an iso-profit graph, where the flexible price is adjusted by two parameters: the price increase ratio $\theta$ and the reservation price ratio $\lambda$. Figure 3.13(a) demonstrates the iso-profit curves under a fixed $\theta$ with value 1.3. The $(L_f, \lambda)$ pairs on each of the solid lines lead to the same expected total profit under the optimal reservation decision, while lines towards the lower-left corner correspond to higher profits than those towards the upper-right corner. Intel can utilize these curves in two ways. The curves demonstrate the dominance between different contract options: for example, contract $A$ with $L_f = 1$, $\lambda = 14\%$, and an expected profit of \$10.950 billion should be preferred to contract $B$ with $L_f = 3$, $\lambda = 12.8\%$, and an expected profit of \$10.893 billion. Alternatively, each curve quantifies the maximum reservation price Intel should be willing to pay for added flexibility; for example, Intel can pay up to 20% of the total price up front and decrease the flexible leadtime to 0 while still keeping its profits at the same level as in contract B. Similarly, Figure 3.13(b) demonstrates the iso-profit curves under a fixed $\lambda$ with value 15%. With the assistance of such iso-profit graphs, Intel will know the bottom-line impact of different alternatives while negotiating with its supplier and will make informed tradeoffs between flexibility and cost. Besides the expected payoff, we also simulate the profit variability associated with different (*leadtime, price*) combinations. For each of the iso-profit curves in the above examples, it




(a) Iso-profit curve when $\theta = 1.3$ (b) Iso-profit curve when $\lambda = 0.15$

Figure 3.13: The Iso-Profit Curves for Fixed $\theta$ and $\lambda$ ($\beta = 0.2,\ \gamma = 3$)

turns out the profit variability first decreases then increases as we move from the left end of the curve to the right. For instance, on the \$10.950B curve in Figure 3.13(a), contract $A$ with $L_f = 1$ leads to the smallest profit c.v. of 17.20%, compared to the $L_f = 0$ case with a c.v. of 17.53% and the $L_f = 3$ case with a c.v. of 18.91%. This implies that a risk-averse firm would prefer a contract term locating in the middle of an iso-profit curve rather than the contracts locating on the two ends.

### 3.4.3 The Impact of Risk Attitude

Since a firm's capacity planning involves a substantial up-front investment with uncertain future revenues, one natural extension of our model is to consider the impact of risk aversion. Van Mieghem (2003) reviews several methods to model a firm's risk aversion and hedging behavior during capacity investment. One predominant approach is to use a concave Bernoulli utility function and assume that the firm operates to maximize its expected *utility* instead of profit. We adopt this approach and inspect how the firm's reservation decisions $B^T$ and $F^T$, as well as the profit and its variability, would change if a concave increasing utility function $\mathcal{G}(\cdot)$ is applied to the reservation stage:

$$\max_{B^T \geq 0; F^T \geq 0} \quad \mathbb{E}_M\ \mathcal{G}(\mathcal{F}(B^T, F^T, M | \vec{\mu}^{1-L_b})),$$

where $\mathcal{F}(B^T, F^T, M|\vec{\mu}^{1-L_b}) = J^{N-L_f}(B^T, F^T, M, \Sigma|\vec{\mu}^{1-L_b}) - r_b B^T - r_f F^T$ and $J^{N-L_f}(\cdot)$ is the value function of the last-stage execution-level optimization problem.

We investigate the case where $\mathcal{G}(\cdot)$ is a power function: $\mathcal{G}(z) = z^\rho$, where $0 < \rho \leq 1$ (see Liu and van Ryzin 2011 for a similar treatment). Note that $\mathcal{G}(z)$ is concave increasing and the smaller the $\rho$, the more risk-averse the firm tends to be (i.e., $\rho = 1$ corresponds to the risk neutrality). Also note that $A(z) = -\frac{\mathcal{G}''(z)}{\mathcal{G}'(z)} = \frac{1-\rho}{z}$ is decreasing in $z$ and $R(z) = zA(z) = 1 - \rho$ is constant in $z$. Hence, $\mathcal{G}(z)$ has decreasing absolute risk aversion (DARA) and constant relative risk aversion (CRRA), two properties that are consistent with experimental and empirical findings about the risk-averse behavior of individuals and corporations (Friend and Blume 1975).

Table 3.2: Risk Aversion Case with Power Utility Function

($\theta = 1.3$, $\lambda = 0.15$, $\beta = 0.2$, $\gamma = 3$, $\psi = 0.95$)

| leadtime | $L_b = 0$ | | | | $L_b = 4,\ L_f = 2$ | | | |
|---|---|---|---|---|---|---|---|---|
| $\rho$ | 1 | 7/8 | 5/8 | 3/8 | 1 | 7/8 | 5/8 | 3/8 |
| $B^T$ | 25.7 | 25.8 | 25.9 | 27.3 | 27.1 | 27.3 | 27.3 | 27.6 |
| $F^T$ | 0 | 0 | 0 | 0 | 3.4 | 3.5 | 3.7 | 3.9 |
| $\frac{F^T}{B^T+F^T}$ | 0% | 0% | 0% | 0% | 11.15% | 11.36% | 11.94% | 12.38% |
| Utility | 1.12 e10 | 6.21 e8 | 1.90 e6 | 5.85 e3 | 1.09 e10 | 6.05 e8 | 1.87 e6 | 5.78 e3 |
| Profit (B) | 11.238 | 11.230 | 11.228 | 11.222 | 10.893 | 10.891 | 10.888 | 10.886 |
| Profit Std | 1.961 | 1.957 | 1.954 | 1.950 | 1.984 | 1.975 | 1.965 | 1.959 |
| Profit c.v. | 17.45% | 17.43% | 17.40% | 17.38 % | 18.21% | 18.13% | 18.04% | 17.98% |

In a numerical analysis (with parameters $\theta$, $\lambda$, $\beta$, and $\gamma$ taking values from a wide range set), we observe that as the firm becomes more risk-averse, it reserves more capacity from both the base and flexible modes. Furthermore, $F^T/(B^T + F^T)$ increases, which implies that the flexible mode becomes a more attractive option. The firm's expected profit decreases due to risk aversion. However, the profit variability (represented by both the standard deviation and c.v.) also decreases. Table 3.2 presents a representative scenario where all the key parameters take their standard values. As a benchmark, we also consider a setting where the base mode has zero leadtime; i.e., the firm enjoys the maximum level of flexibility and cost efficiency. Even under this ideal setting, the coefficient of variation of the total profit is around 17%. This level is inevitable due to the demand forecast evolution process and the tactical nature of the problem. That is, risk aversion removes a limited portion of the profit variability arising from operational demand-supply mismatches.

## 3.5. Conclusion

Capital equipment purchasing is a crucial yet difficult task for many semiconductor, electronic, automotive, and pharmaceutical firms. In this chapter, we proposed a dual-mode equipment procurement model (DMEP) to guide firms through this complex task. DMEP serves three roles. On the strategic level, it provides decision support to contract negotiation by comparing alternatives with different levels of flexibility and costs. On the tactical level, it guides capacity reservation decisions by characterizing the amount of capacity that should be reserved from different procurement modes. On the operational level, it quantifies procurement amounts by considering the latest demand information as well as the installed capacity. By incorporating interactions between these three levels of decision-making, DMEP enables a flexible supply chain that adapts effectively to changing demand conditions. It helps firms better manage their equipment procurement process, eliminate excess capacity, and thus lower their costs. It benefits suppliers by enabling a risk-sharing mechanism through up-front capacity reservation and the elimination of soft orders.

DMEP formalizes and extends the approach that Intel has used in the past to price and exercise capacity options with reduced leadtimes for a few types of fabrication equipment (Vaidyanathan *et al.* 2005). It also complements recent improvements in demand forecasting methodologies at Intel (Wu *et al.* 2010). The dual-mode procurement approach described here is being used at the strategic and tactical levels today at Intel for all types of fabrication equipment, and will soon be used at the operational level. Implementation of the approach has leveraged model structure and details to provide the types of sensitivity analysis needed by Intel decision-makers to understand and take advantage of the subtleties of this improvement in the capital equipment acquisition process. As demonstrated by Table 3.5, the annual savings on capital procurement at Intel due to implementing DMEP are estimated to exceed tens of millions of dollars.

Table 3.3: Comparison of DMEP and the Default Single-Sourcing Approach at Intel (For the numerical analysis, unless otherwise stated, $\theta = 1.3$, $\lambda = 0.15$, $\beta = 0.2$, $\gamma = 3$, and $\psi = 0.95$)

| | | | | | | |
|---|---|---|---|---|---|---|
| **The impact of flexible cost increase ratio $\theta$** | | | | | | |
| $\theta$ | 1.2 | 1.3 | 1.4 | 1.5 | 1.6 | |
| Expected Profit with DMEP (in billions) | 10.905 | 10.893 | 10.881 | 10.872 | 10.861 | |
| Expected Profit with Default Approach (approximation, in billions) | 10.802 | 10.802 | 10.802 | 10.802 | 10.802 | |
| Improvement | 0.95% | 0.84% | 0.73% | 0.65% | 0.55% | |
| **The impact of flexible reservation ratio $\lambda$** | | | | | | |
| $\lambda$ | 0.05 | 0.10 | 0.15 | 0.20 | 0.25 | |
| Expected Profit with DMEP (in billions) | 11.036 | 10.957 | 10.893 | 10.847 | 10.812 | |
| Expected Profit with Default Approach (approximation, in billions) | 10.802 | 10.802 | 10.802 | 10.802 | 10.802 | |
| Improvement | 2.17% | 1.43% | 0.84% | 0.42% | 0.09% | |
| **The impact of flexible mode leadtime $L_f$** | | | | | | |
| $L_f$ | 0 | 1 | 2 | 3 | 4 | |
| Expected Profit with DMEP (in billions) | 11.001 | 10.918 | 10.893 | 10.847 | 10.802 | |
| Expected Profit with Default Approach (approximation, in billions) | 10.802 | 10.802 | 10.802 | 10.802 | 10.802 | |
| Improvement | 1.84% | 1.07% | 0.84% | 0.42% | 0.00% | |
| **The impact of $\gamma$** | | | | | | |
| $\gamma$ | 0 | 1 | 2 | 3 | 4 | 5 |
| Expected Profit with DMEP (in billions) | 11.09 | 11.031 | 10.965 | 10.893 | 10.817 | 10.742 |
| Expected Profit with Default Approach (approximation, in billions) | 10.993 | 10.904 | 10.807 | 10.802 | 10.603 | 10.495 |
| Improvement | 0.88% | 1.16% | 1.46% | 0.84% | 2.02% | 2.35% |
| **The impact of mean jump size $\beta$** | | | | | | |
| $\beta$ | 0% | 10% | 20% | 30% | 40% | 50% |
| Expected Profit with DMEP (in billions) | 11.998 | 11.574 | 10.893 | 10.023 | 8.989 | 7.786 |
| Expected Profit with Default Approach (approximation, in billions) | 11.905 | 11.446 | 10.802 | 9.755 | 8.611 | 7.271 |
| Improvement | 0.78% | 1.12% | 0.84% | 2.75% | 4.39% | 7.08% |

Despite its many advantages, DMEP also has several limitations. First, we investigated a situation in which the firm procures only one type of equipment from its supplier. This simplification enabled us to demonstrate the dynamics of the algorithm without introducing complexity. The DMEP model can be generalized to a multi-equipment scenario in which (1) the firm considers ordering from the base and flexible modes for all types of equipment with different leadtimes in each period and (2) the available capacity in each period is constrained by the lowest capacity among all the types of equipment. As expected, as the number of types of equipment increases, the interactions and the complexity of the problem also increase. Having said that, firms should consider DMEP only for equipment on the critical path of capacity planning as the rest (in Intel's case the types that have shorter leadtimes and are cheaper) will not impose additional constraints on the system. Yet, alternative formulations for multi-equipment procurement would be valuable. These alternative formulations should consider the multi-tool problem as a portfolio of tools, suppliers, and potentially approaches other than DMEP. Second, we formulated the procurement problem as a linear program. As such, DMEP is better suited to providing the fraction of orders from each mode rather than the actual procurement quantities. However, if the intention is to use DMEP to obtain the specific order quantities, generalizing the DMEP model to an integer program would be preferable. Finally, we ignored the inventory level decision of the firm and assumed a build-to-order system. At the top level of the framework we are solving a strategic problem possibly years in advance of actual demand realization. Given this long leadtime, there is huge uncertainty regarding not only the demand but also the supply process (e.g., yields). Therefore, we simplified the execution-level problem and concentrated on the most critical decisions (that is, the capacity levels to be executed) to avoid including additional noise. The decision to not include product inventory is also consistent with Intel's strategy of erring on the side of ordering too much equipment, and is considered to be a better initial approach than tackling too many frontiers at the same time. If we were to include product inventories, we would see reductions in the total capacity level as capacity and inventory are substitutes. We would like to note that the execution-level algorithm can be modified to include the option of holding product inventory easily. One must define a decision variable for product inventory for each period and parameters for inventory holding cost and salvage value. As a result, the execution-level problem would be slightly more complicated since, in addition to the base and flexible capacity execution levels, the optimization will also need to calculate the optimal product inventory levels as well.

This chapter took an initial, yet important, step towards formulating the complex dual-mode equipment procurement problem in a practical and insightful way. In capital-intensive industries, capital expenditures often constitute about one quarter of the total revenue and roughly two thirds of the manufacturing costs. Given the millions of dollars that are at stake, we believe that equipment procurement problems will attract more attention from academia. Fortunately, there is plenty of room for interesting future work around this topic.

# Chapter 4

# Strategic Capacity Allocation in Commodity Trading with a Spot Market

## 4.1. Introduction

As the key ingredient for the world's most commonly used metal – steel, which represents almost 95% of all metal used per year (Blas 2009), iron ore has always been one of the most important commodities traded in the global market. Christopher LaFemina, mining analyst at Barclays Capital, once said that "Iron ore may be more integral to the global economy than any other commodity, except perhaps oil (Blas 2009)." For the past decade, due to the rapid development of China and other Asian countries, the already-enormous iron ore market has been growing at 10% per annum on average; and in 2010, the seaborne iron ore market, that is, iron ore to be shipped to other countries across the ocean, reached a historical revenue size of $88 billion (Serapio and Trevethan 2010). Given such a large amount of capital at stake, players in this industry, especially the iron ore producers, must be extremely cautious with their global operational strategies, such as capacity planning, pricing, distributing, etc., since even a small misplay could lead to hundreds of millions of dollars' loss of revenue as well as a potentially unfavorable position in relation to the competition.

Because of the intensive investment in capital equipment and transportation infrastructure that is required at the initial stage of the mining operations, global iron ore production has been concentrated in the hands of a few major players. The world's three largest iron ore producers, Vale S.A. from Brazil, Rio Tinto and BHP Billiton from Australia, control over two-thirds of the world's iron ore supply and jointly account for more than 23% of the total assets of the Market Vectors Steel ETF (Ausick 2010). In the past, the iron ore business for these big suppliers had been straightforward: Their clients were mostly state-owned steel manufacturers with blast furnaces, such as Nippon Steel of Japan, POSCO of Korea, US Steel, and BaoSteel of China. These steel mills had largely predictable business and preferred long-term relationships with the iron ore suppliers; as a result, the Big Three sold almost all of their iron ore through forward contracts at a pre-negotiated price, which was called the World Benchmark Price (Lee 2007). In this relatively stable environment, the key to iron ore suppliers increasing their profits lies in their ability to negotiate the forward price and to reduce the operating costs.

However, changes to the system have been occurring in recent years. As China's economy continues to grow by double digits, its demand for steel, and consequently iron ore, soared along with the booming infrastructure constructions throughout the country. For instance, China imported 275 million metric tons of iron ore in 2005, and that number jumped to around 320 million in 2006, suggesting a 16.4% increase (Lee 2007). As a result, the delivering capacity negotiated through forward contracts between the iron ore producers and the steelmakers has often proved to be insufficient, creating certain spot markets where some local small iron ore suppliers filled the gap by trading with the steelmakers at the then-current spot price, which is usually higher than the fixed contract price. This phenomenon implies a potential revenue loss for the Big Three since they have committed almost all of their capacity to the contract channel and hence cannot benefit from the high spot price. For instance, in 2005, nearly half of the 275 million metric tons of iron ore imported to China were purchased based on spot prices; while the Big Three together only supplied 25 million metric tons through the spot market. Hence, it is no longer optimal for the big iron ore suppliers to only use forward contracts to deploy their capacity.

Given this emerging situation, the big iron ore suppliers are faced with at least two challenges: first, constrained capacity that limits the firms' ability to meet the continuous demand surge in China and other Asian countries; second, the need for an efficient strategy to properly allocate sales quantities to both the contract channel and the spot channel in order to achieve a tradeoff between stable production from the former and a potentially higher profit margin from the latter. The solution to the first challenge is of course expansion, which for the mining industry usually implies hundreds of millions of dollars' investment and several years' constructing operations. Rio Tinto, for example, has committed about $6 billion of new investment in the Pilbara region of Western Australia since 2010, with the majority to be spent on expansion.[1] Although crucial, this strategic level decision is not overly complex to make, since much randomness can be ignored due to the pooling effect over such a long planning horizon.

The second challenge, which lies more on the tactical level with a planning horizon of quarters, is nevertheless trickier to tackle. On one hand, the business environment contains several random factors, such as the total contract channel demand and the equilibrium spot price, both of which contribute to a stochastic planning problem. On the other hand, the decisions are intertwined since the capacity that the iron ore supplier allocates to the contract channel may finally affect the spot price through impacting the demand and supply curves in the spot market. One recent strategic move of the Big Three in terms of utilizing the spot price was to shift the long-term contract[2] mechanism from an annual-review basis into a quarterly-review basis and to set the contract price based on the previous period's average spot price. "This change has come from the suppliers' desire to see prices more closely mirror the spot market ⋯ following several years in which spot prices have exceeded long-term contract prices (Burns 2010)." Meanwhile, the big suppliers also expect to significantly strengthen their presence in the spot market. Graeme Stanway, former chief iron-ore consultant at Rio Tinto, told the authors in November 2010 that "spot tonnes had only previously made up a small part of the sales portfolio (of Rio Tinto) but would be gradually increased to 50%." However, such decisions are not necessarily established based on rigorous quantitative analysis.

How the big iron ore suppliers should manage the aforementioned two selling

[1] http://www.riotintoironore.com/ENG/operations/301_expansion_projects.asp

[2] In this chapter we use *forward contract* and *long-term contract* interchangeably.

channels by strategically allocating capacity between the forward contract and the spot market is the focus of this research. Due to the complexity of the actual business setting, we need to make two simplifying assumptions here: 1. We eliminate the concern about competition by treating the Big Three as one company and calling it the "Supplier." 2. We start with a single period model in which the forward contract price is fixed according to the previous period's realized spot price and hence the supplier only has quantity decisions to make; however, there can be multiple decision stages within that single period. We then investigate several versions of the problem with different ways to model the spot market to generate far-ranging insights with diverse practical emphasis. More specifically, we first study a case in which the spot market is *open*, i.e., the spot price is an exogenous random variable not affected by the players' actions. We show that if the contract channel demand is unlimited, then the supplier will adopt a non-extreme policy, i.e., allocating part of the capacity to the contract channel and the rest to the spot channel, only if he is risk-averse. When the contract channel demand is stochastic and satisfies a bivariate normal distribution with the spot price, however, the supplier's expected profit function is concave-convex and an interior optimal policy may exist even when the supplier is risk-neutral. Furthermore, the supplier tends to allocate more quantity to the spot channel if the contract channel demand and the spot price are more positively correlated, and he should allocate more to the contract channel if he is more risk-averse. We then investigate the case where the spot market is "closed," i.e., the equilibrium spot price is endogenously determined by the spot demand curve and the spot supply curve, both of which are affected by the supplier's allocation decision. We demonstrate that if the shifting effect of the supplier's second stage quantity decision on the default spot supply curve is stronger than the shifting effect of the unfulfilled first stage demand on the default spot demand curve, then the supplier's first stage expected profit function is convex-concave and an interior optimal solution may exist; otherwise, an extreme policy would be optimal in the first stage. We perform numerical analysis to gauge the sensitivity of the parameters and generate additional managerial insights. It should be emphasized that although this research is motivated by the iron ore industry, the modeling approach and managerial insights should be applicable and illuminative to other similar commodity trading businesses as well.

## 4.2. Literature Review

Two bodies of literature are most related to this research: distribution channel management and the dynamics of spot markets and forward contracts.

The multi-channel strategy in the context of procurement management has been well studied for decades. Recent work includes Minner (2003), who provides a thorough review of the multi-source inventory control models, and Chao *et al.* (2009) and Peng *et al.* (2010), who investigate multi-sourcing capacity expansion problems with lost sales and backorders, respectively. In the recent decade, multi-channel distribution strategies also received significant attention from both industry and academia due to the prevalence of e-commerce and other internet-enabled opportunities. Chiang *et al.* (2003) identify how a manufacturer may open a direct selling channel to compete with its own retailers if the customer acceptance of the direct channel is strong enough. Reinhardt and Levesque (2004) use microeconomics to study how a firm should allocate its sales quantity between an online direct channel and an offline channel to best trade off cost, revenue, and competitive behavior; they demonstrate that it may not be optimal for the firm to sell in both channels. Chen *et al.* (2008) introduce a consumer channel choice model and gauge its impact on the manufacturer's choice between a direct channel and a traditional retail channel. Tsay and Agrawal (2004) provide a comprehensive review on the modeling of multi-channel distribution systems. In this chapter, the supplier's allocation of production capacity is essentially a channel management problem. However, instead of the online and offline channels, we investigate the supplier's choice between a contract channel and a spot market channel; moreover, in our model the supplier is dealing with manufacturers rather than end consumers.

In relation to modeling the spot market, up to now, most researchers have chosen to work with an *open* spot market, where the spot price is independent of the actions of individual market participants. A typical model treats the spot price as a stochastic variable, the distribution of which is known to the decision-maker (Wu *et al.* 2002, Golovachkina and Bradley 2002, Seifert *et al.* 2004, etc.). A more elegant approach to model a multi-period problem is to assume that the spot price follows a Markovian stochastic process (Kalymon 1970, Assuncao and Myer 1993, Secomandi 2010) or a geometric Brownian motion (Dixit and Pindyck 1994, Li and Kouvelis 1999), the former of which applies to the discrete case and the latter the continuous case. In the first half of this chapter, we examine a preliminary model where the spot market is open and the spot price follows a bivariate normal distribution with the contract channel demand. We model it in a way similar to Seifert *et al.* (2004). The difference is that in their model a manufacturer uses the spot market as a backup sourcing channel, whereas in our problem a capacitated supplier treats the spot market as an extra distribution channel.

Some researchers have attempted to model a *closed* spot market, where the actions of market participants can affect the price. A typical approach is to identify the spot market price using the rational expectations equilibrium approach (Grossman 1981, Kyle 1989, etc.) – at the equilibrium, the total supply quantity derived from the sellers' supply curves must equal the total demand quantity obtained according to the buyers' demand curves. Lee and Whang (2002) utilize this concept and examine the spot trading of excess inventory in terms of a secondary market, where the equilibrium price is endogenously determined. They demonstrate that with a larger number of buyers, the secondary market can increase the allocation efficiency of the supply chain, but not necessarily the sales of the manufacturer. Kleindorfer and Wu (2003) integrate long-term contracting with spot trading via B2B exchanges for capital-intensive industries. They use a general framework based on transaction cost economics to provide a synthesis of the existing literature. Some researchers directly make spot price the supplier's decision variable. Erhun *et al.* (2000), for example, look at a decentralized supply chain where a manufacturer procures capacity from a single supplier through a spot market over multiple periods, where the spot price is set by the supplier. They show that double marginalization can be reduced or even entirely eliminated by increasing the number of trading periods. In the second half of this chapter, we focus on modeling the formation of the spot price endogenously, as a consequence of the equilibrium outcome. A unique feature of our model is that the supplier's single quantity decision can affect both the demand curve and the supply curve in the spot market, since part of the unfulfilled demand from the contract channel will later switch to the spot market. Haksoz and Seshadri (2007) carry out a complete survey on the use of spot markets to manage procurement in supply chains.

There has also been some work that discusses the employment of both a forward (fixed-price) contract and a spot market as distribution channels. Allaz (1992) builds a two-period model of an oligopoly producing a homogeneous good which is traded first on a forward market and then on a spot market. He shows that forward transactions can be used strategically by the producers to improve their positions in the spot market. Liski and Montero (2006) investigate an infinitely repeated oligopoly in which firms participate in both the spot market and forward transactions. They demonstrate that forward trading enables firms to achieve collusive profits. These two works simplify the modeling of the spot price by either introducing an exogenous random variable or assuming a linear inverse demand curve. In another closely related work, Mendelson and Tunca (2007) investigate a dynamic supply chain trading game between a supplier and several buyers who first sign fixed-price contracts and then trade in the spot market once the private information of each party is revealed. In this scenario, the spot price is determined based on rational expectations equilibrium. They find that while spot trading helps reduce prices, increase the produced quantity, and improve supply chain profits, it does not eliminate the fixed-price contracting. In contrast to our work, however, none of the above papers assumes a constrained capacity for the supplier, which is indeed the case with iron ore producers; nor is there a connection between the contract channel demand and the spot market demand.

The rest of this chapter is organized as follows: In Section 4.3, we introduce the basic setting of the business problem and define the key parameters of the model. In Section 4.4, we discuss the supplier's capacity allocation strategy under an open spot market, where the equilibrium spot price is given by an exogenous random variable. Section 4.5 extends the investigation to a closed spot market scenario in which the equilibrium spot price is determined based on the demand and supply curves in the spot market, both of which are affected by the supplier's capacity allocation decision. Section 4.5.4 presents numerical analysis. Section 4.6 concludes the chapter and delivers managerial insights.

## 4.3. The Business Setting

Figure 4.1 demonstrates the basic setting of the business problem. Consider a commodity trading supply chain in which an oligopoly Supplier with total capacity $K$

(within a certain period) determines how to allocate his potential sales quantity between a forward contract channel and a spot market channel to maximize his total expected profit. The contract channel has a given price $w$, which in reality is fixed according to the previous period average spot price and hence is not a decision variable. The spot price $p_s$ is either an exogenous random variable with a known distribution, or is to be endogenously determined by the demand and supply curves. A group of large buyers with aggregated stochastic demand $D$ treats the forward contract channel with a higher priority, that is, $D$ is first satisfied via the contract channel. This happens in practice because large steel manufacturers with high volume prefer a stable iron ore price to minimize procurement cost volatility. If the quantity $\Lambda$ that the supplier allocated to the contract channel is insufficient, however, part of the unfulfilled demand $(D-\Lambda)^+$ will then switch to the spot market. Depending on whether the spot price is exogenous or endogenous, the supplier will allocate either all or part of his leftover capacity $K - \min(\Lambda, D)$ to the spot market for sale. Note that besides the oligopoly supplier and the big buyers, for the most comprehensive model we assume there are some small local suppliers and buyers dealing in the spot market as well, leading to the default spot supply and demand curves.

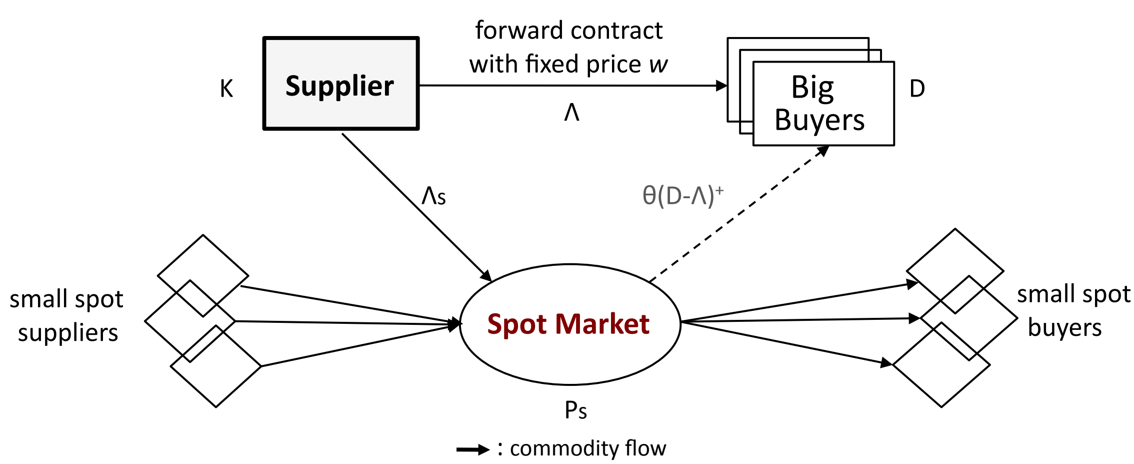



Figure 4.1: Commodity Trading Participants and Their Business Relationships

## 4.4. Strategic Allocation under an Open Spot Market

Before investigating the comprehensive case where the commodity supplier's quantity decision can affect the equilibrium spot price, we first discuss a simpler setting in which the spot market is open and the equilibrium spot price is described by an exogenous random variable. This discussion can be instructive to small- or medium-sized commodity suppliers that do not have large enough market influence to impact the formation of the spot price. In a single period problem, a commodity supplier $S$ with a total capacity $K$ tries to divide the capacity between two selling channels: a contract channel with a fixed price $w$ and a spot market with a random price $p_s$, which has a p.d.f. of $\phi_s(\cdot)$, a mean of $\mu_s$, and a standard deviation of $\sigma_s$. The allocation decision is made before the randomness of the spot price is resolved. We assume a unit production cost of $c$ that is lower than both $w$ and $p_s$.

### 4.4.1 Unlimited Contract Channel Demand

If the supplier has a tight capacity that is with certainty lower than the potential contract channel demand, i.e., the contract demand can be treated as unlimited, then the risk-neutral supplier's optimal expected profit is given by $\pi_S = \max_{0 \le \Lambda \le K} \mathbb{E}_{p_s}[w\Lambda + p_s(K - \Lambda) - cK]$, where the decision variable $\Lambda$ represents the amount of capacity allocated to the contract channel. It is trivial to verify that the supplier's optimal policy under this scenario is of an extreme type: if $\mu_s \le w$, then $\Lambda^* = K$; if $\mu_s > w$, then $\Lambda^* = 0$.

Now, assume the supplier is risk-averse and operates to maximize his mean-variance utility. That is,

$$\pi_S = \max_{0 \le \Lambda \le K} \mathbb{E}\Pi_S(\Lambda) - kVar\Pi_S(\Lambda), \tag{4.4.1}$$

where $\Pi_S(\Lambda) = w\Lambda + p_s(K - \Lambda) - cK$, and $k > 0$ represents the supplier's risk-averse magnitude. His optimal allocation policy is described in the following proposition:

**Proposition 4.4.1.** *Under risk aversion, the supplier's optimal allocation decision*

$\Lambda^*$ *is given by:*

$$\Lambda^* = \begin{cases} K, & \text{if } \mu_s \in (0,\ w] \\ K - \dfrac{\mu_s - w}{2k\sigma_s^2}, & \text{if } \mu_s \in (w,\ w + 2k\sigma_s^2 K] \\ 0, & \text{if } \mu_s > w + 2k\sigma_s^2 K \end{cases} \tag{4.4.2}$$

Proposition 4.4.1 says that when the supplier is risk-averse, he may adopt a mixed portfolio by selling through both the contract channel and spot channel if the expected spot price is higher than the fixed contract price but not too high. The more risk-averse the supplier is, the more quantity he should allocate to the contract channel.

### 4.4.2 Stochastic Contract Channel Demand

If the supplier's total capacity is relatively high, though, it is more reasonable to also treat the contract channel demand as a random variable. Here we adopt the idea of Seifert *et al.* (2004) and assume that the spot price $p_s$ and the contract channel demand $D$ follow a bivariate normal distribution, i.e., $(p_s, D) \sim BN(\mu_s, \mu_d, \sigma_s^2, \sigma_d^2, \rho)$. Let $\phi_{s,d}(\cdot)$ be the joint density function of the bivariate normal distribution and let $\phi_d(\cdot)$ represent the p.d.f. of the normal distribution $N(\mu_d, \sigma_d^2)$. $\rho > 0$ implies a positive correlation between $p_s$ and $D$; this is usually the case since a high contract channel demand likely suggests the popularity of the commodity and hence a high spot price as well. In contrast, $\rho < 0$ means $p_s$ and $D$ are negatively correlated; this could also be the case if there is limited total demand in the market, and thus a high contract channel demand would imply a relatively lower spot demand, leading to a lower spot price. The sequence of events within a period is: First, the supplier allocates $\Lambda \in [0, K]$ to the contract channel. Then, both $D$ and $p_s$ are realized. Last, the supplier sells the leftover capacity $K - \min(\Lambda, D)$ in the spot market.

We start by investigating a risk-neutral case. The supplier's optimal expected profit is given by $\pi_S = \max\limits_{0 \le \Lambda \le K} \mathbb{E}_{p_s,D}\, \Pi_S(\Lambda)$, where

$$\begin{aligned} \Pi_S(\Lambda) &= w \min(\Lambda, D) + p_s(K - \min(\Lambda, D)) - cK \\ &= (w - p_s)\min(\Lambda, D) + (p_s - c)K. \end{aligned} \tag{4.4.3}$$

More explicitly, the objective function can be written as[3]

$$
\begin{aligned}
\mathbb{E}\Pi_S(\Lambda) &= \int_{x=0}^{\Lambda} x \int_{p_s=0}^{\infty} (w-p_s)\phi_{s,d}(p_s,x)dp_s dx \\
&\quad + \int_{x=\Lambda}^{\infty} \Lambda \int_{p_s=0}^{\infty} (w-p_s)\phi_{s,d}(p_s,x)dp_s dx + (\mu_s - c)K \\
&= \int_{x=0}^{\Lambda} x[w-(\mu_s+\rho\frac{\sigma_s}{\sigma_d}(x-\mu_d))]\phi_d(x)dx \\
&\quad +\Lambda \int_{x=\Lambda}^{\infty} [w-(\mu_s+\rho\frac{\sigma_s}{\sigma_d}(x-\mu_d))]\phi_d(x)dx + (\mu_s-c)K \\
&= (w-\mu_s+\rho\frac{\sigma_s}{\sigma_d}\mu_d)[\Lambda(1-\Phi_d(\Lambda)) + \int_0^{\Lambda} x\phi_d(x)dx] \\
&\quad -\rho\frac{\sigma_s}{\sigma_d}[\Lambda \int_{\Lambda}^{\infty} x\phi_d(x)dx + \int_0^{\Lambda} x^2\phi_d(x)dx] + (\mu_s-c)K. \qquad (4.4.4)
\end{aligned}
$$

Proposition 4.4.2 below captures the supplier's optimal allocation strategy.

**Proposition 4.4.2.** *Assume* $\mu_d$ *is large enough so that* $\phi_d(0) \approx 0$*; we then have:*
*If* $\rho \geq 0$*, then* $\mathbb{E}\Pi_S(\Lambda)$ *is concave for* $\Lambda \in [0, \max(0, \min(K, \mu_d + \frac{(w-\mu_s)\sigma_d}{\rho\sigma_s}))]$ *and convex decreasing for* $\Lambda \in (\max(0, \min(K, \mu_d + \frac{(w-\mu_s)\sigma_d}{\rho\sigma_s})), K]$*. The supplier's optimal allocation decision* $\Lambda^*$ *is given by*

$$
\Lambda^* = \begin{cases} \min(K, \hat{\Lambda}), & \text{if } \mu_s \leq w \\ 0, & \text{if } \mu_s > w \end{cases} \qquad (4.4.5)
$$

*where* $\hat{\Lambda}$ *satisfies* $(w-\mu_s+\rho\frac{\sigma_s}{\sigma_d}\mu_d)(1-\Phi_d(\hat{\Lambda})) - \rho\frac{\sigma_s}{\sigma_d}\int_{\hat{\Lambda}}^{\infty} x\phi_d(x)dx = 0.$
*If* $\rho < 0$*, then* $\mathbb{E}\Pi_S(\Lambda)$ *is convex for* $\Lambda \in [0, \max(0, \min(K, \mu_d + \frac{(w-\mu_s)\sigma_d}{\rho\sigma_s}))]$ *and concave increasing for* $\Lambda \in (\max(0, \min(K, \mu_d + \frac{(w-\mu_s)\sigma_d}{\rho\sigma_s})), K]$*. The supplier's optimal allocation decision* $\Lambda^*$ *is of an extreme type and given by:*

$$
\Lambda^* = \begin{cases} K, & \text{if } \mu_s \leq w; \text{ or } \mu_s > w,\ \mathbb{E}\Pi_S(K) \geq (\mu_s - c)K \\ 0, & \text{if } \mu_s > w,\ \mathbb{E}\Pi_S(K) < (\mu_s - c)K \end{cases} \qquad (4.4.6)
$$

We see from above that when the contract demand $D$ and the spot price $p_s$ are

[3] If random variables $X$ and $Y$ satisfy a bivariate normal distribution $BN(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, then the conditional distribution $Y|X \sim N(\mu_Y + \frac{\sigma_Y}{\sigma_X}\rho(X-\mu_X), (1-\rho^2)\sigma_Y^2)$.

positively correlated, i.e., $\rho \geq 0$, the supplier may only allocate part of his capacity to the contract channel even if the expected spot price $\mu_s$ is lower than the fixed contract price $w$. This result is in stark contrast with the unlimited contract demand scenario, in which case the optimal policy is of an extreme type – either allocating all capacity to the contract channel or allocating all to the spot channel, and an interior optimizer may exist only if the supplier is risk-averse. Below we provide an explanation for why an interior optimizer exists for the risk-neutral case here: Since we always allow the supplier to relocate his unused capacity $(\Lambda - D)^+$ from the contract channel back to the spot channel, the only circumstance that may penalize the supplier for over-allocating quantity to the contract channel would be that both the realized contract demand and the realized spot price are high (thus the supplier wouldn't have extra capacity to benefit from the high spot price), which is only likely to happen if $\rho$ is positive. When $D$ and $p_s$ are negatively correlated, i.e., $\rho < 0$, though, the policy tends to be extreme again. However, even when the expected spot price $\mu_s$ is higher than the contract price $w$, the supplier may still preallocate all quantity to the contract channel. The reason is that this time a high contract channel demand would signal a low spot market price, and hence it is "safe" to exploit the contract channel first. The following proposition focuses on the $\rho \geq 0$ case that is more relevant in practice and discusses some comparative statics results with respect to the interior maximizer $\hat{\Lambda}$. More detailed analysis on the monotonicity of the optimal decision in the actual business setting will be implemented numerically in Section 4.4.3.

**Proposition 4.4.3.** *(Comparative Statics) The interior maximizer $\hat{\Lambda}$ is increasing in $w$ and $\mu_d$, and decreasing in $\mu_s$. In addition, if $\hat{\Lambda} \geq \mu_d$, then $\hat{\Lambda}$ is increasing in $\sigma_d$, and decreasing in $\sigma_s$ and $\rho$.*

Next, we discuss the impact of the supplier's risk attitude on his optimal capacity allocation decision. We still adopt a mean-variance approach; hence, when the commodity supplier is risk-averse, he solves the following optimization problem:

$$\pi_S = \max_{0 \leq \Lambda \leq K} \mathbb{E}_{p_s,D}\Pi_S(\Lambda) - kVar\Pi_S(\Lambda), \tag{4.4.7}$$

where $\Pi_S(\Lambda) = (w - p_s)\min(\Lambda, D) + (p_s - c)K = (w - p_s)[\Lambda - (\Lambda - D)^+] + (p_s - c)K$.

We have already explicitly described $\mathbb{E}\Pi_S(\Lambda)$ in Equation (4.4.4). Equation (4.4.8) and (4.4.9) below further capture $Var\Pi_S(\Lambda)$ and its first order derivative with respect to $\Lambda$. Unfortunately, the equations are too complex and one needs to resort to numerical methods for a solution. On a high level, let $\hat{\Lambda}$ denote the solution to the first order condition $\frac{\partial \mathbb{E}\Pi_S(\Lambda)}{\partial \Lambda} - k\frac{\partial Var\Pi_S(\Lambda)}{\partial \Lambda} = 0$; then the optimal allocation decision should be one of 0, $K$, and $\hat{\Lambda}$, whichever leads to the highest utility.

$$\begin{aligned}
&Var\Pi_S(\Lambda) \\
&= (1-\rho^2)\sigma_s^2[\Lambda^2 + \int_0^\Lambda (\Lambda - x)^2\phi_d(x)dx - 2\Lambda\mathbb{E}(\Lambda - D)^+ + K^2] \\
&+(K-\Lambda)^2\rho^2\sigma_s^2 + \int_0^\Lambda (w - \mu_s - \rho\frac{\sigma_s}{\sigma_d}(x-\mu_d))^2(\Lambda - x)^2\phi_d(x)dx \\
&+2(K-\Lambda)\rho\frac{\sigma_s}{\sigma_d}\mu_d\int_0^\Lambda (w - \mu_s - \rho\frac{\sigma_s}{\sigma_d}(x-\mu_d))(\Lambda - x)\phi_d(x)dx \\
&-2(K-\Lambda)\rho\frac{\sigma_s}{\sigma_d}\int_0^\Lambda (w - \mu_s - \rho\frac{\sigma_s}{\sigma_d}(x-\mu_d))x(\Lambda - x)\phi_d(x)dx \\
&-[\int_0^\Lambda (w - \mu_s - \rho\frac{\sigma_s}{\sigma_d}(x-\mu_d))(\Lambda - x)\phi_d(x)dx]^2.
\end{aligned} \tag{4.4.8}$$

$$\begin{aligned}
&\frac{\partial Var\Pi_S(\Lambda)}{\partial \Lambda} \\
&= 2\sigma_s^2\Lambda - 2(1-\rho^2)\sigma_s^2\Lambda\Phi_d(\Lambda) - 2K\rho^2\sigma_s^2 \\
&+2(K-2\Lambda)\rho\frac{\sigma_s}{\sigma_d}\mu_d\int_0^\Lambda (w - \mu_s - \rho\frac{\sigma_s}{\sigma_d}(x-\mu_d))\phi_d(x)dx \\
&+2(2\Lambda - K + \mu_d)\rho\frac{\sigma_s}{\sigma_d}\int_0^\Lambda (w - \mu_s - \rho\frac{\sigma_s}{\sigma_d}(x-\mu_d))x\phi_d(x)dx \\
&-2\rho\frac{\sigma_s}{\sigma_d}\int_0^\Lambda (w - \mu_s - \rho\frac{\sigma_s}{\sigma_d}(x-\mu_d))x^2\phi_d(x)dx \\
&+2\int_0^\Lambda (w - \mu_s - \rho\frac{\sigma_s}{\sigma_d}(x-\mu_d))^2(\Lambda - x)\phi_d(x)dx \\
&-2\int_0^\Lambda (w - \mu_s - \rho\frac{\sigma_s}{\sigma_d}(x-\mu_d))(\Lambda - x)\phi_d(x)dx\int_0^\Lambda (w - \mu_s - \rho\frac{\sigma_s}{\sigma_d}(x-\mu_d))\phi_d(x)dx.
\end{aligned} \tag{4.4.9}$$

### 4.4.3 Numerical Analysis

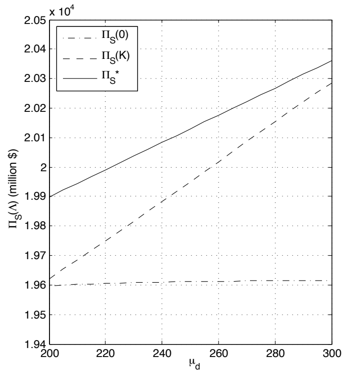
In this part we concentrate on the case where the contract channel demand $D$ and the spot market price $p_s$ follow a bivariate normal distribution. Table 4.1 below summarizes the benchmark values for all model parameters, which are based on practical estimates and largely consistent with the future numerical study in Section 4.5.4. We investigate how the supplier's optimal allocation decision and the corresponding expected profit would change with respect to six parameters: $\mu_d$, $\mu_s$, $\sigma_d$, $\sigma_s$, $\rho$, and $k$, whose impacts have not been fully characterized by Proposition 4.4.3. Unless otherwise stated, for all numerical examples we vary only one target parameter while fixing the remaining parameters at their benchmark values.

Table 4.1: Benchmark Values of Model Parameters (units: $K, \mu_d, \sigma_d$: million ton; $w, c, \mu_s, \sigma_s$: dollar per ton; $k, \rho$: no unit)

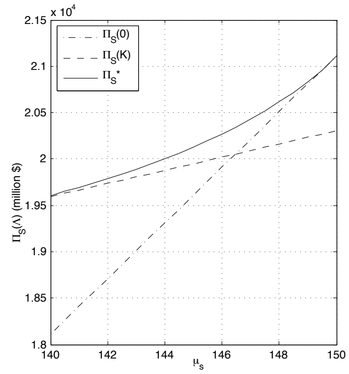
| parameter | value | parameter | value | parameter | value |
|---|---|---|---|---|---|
| $K$ | 300 | $w$ | 150 | $c$ | 80 |
| $k$ | 0 | $\mu_d$ | 250 | $\sigma_d$ | 100 |
| $\rho$ | 0.5 | $\mu_s$ | 145 | $\sigma_s$ | 20 |

#### The Impact of $\mu_d$ and $\mu_s$

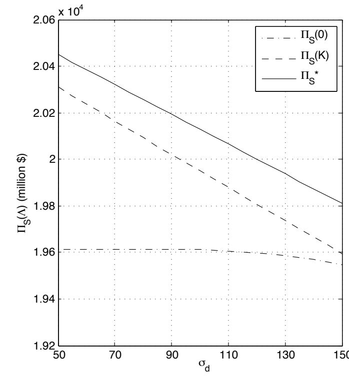
We refer to the policy of allocating all capacity to the contract channel (i.e., $\Lambda = K$) as the *total-contract* policy, and the policy of allocating all the capacity to the spot market (i.e., $\Lambda = 0$) as the *total-spot* policy. Figure 4.2(a) and 4.2(b) demonstrate that: The supplier's optimal expected profit is increasing in the average contract channel demand $\mu_d$ and the average spot price $\mu_s$. The profit gap between the optimal allocation strategy (with a mixed portfolio) and the total-contract strategy is decreasing in $\mu_d$ and increasing in $\mu_s$, with average profit improvements of 0.89% and 1.28%, respectively. The profit gap between the optimal allocation strategy and the total-spot strategy is increasing in $\mu_d$ and decreasing in $\mu_s$, with average profit improvements of 2.65% and 3.15%, respectively. Finally, Figure 4.3(a) shows that the optimal contract channel allocation quantity $\Lambda^*$ is increasing in $\mu_d$ and decreasing in $\mu_s$, which is consistent with the result in Proposition 4.4.3.



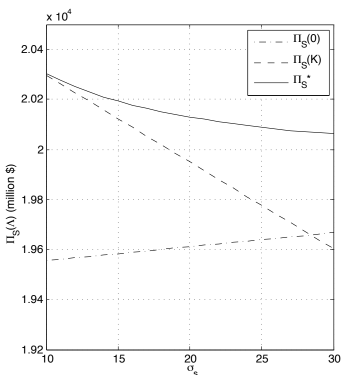

(a) Profit Comparison with Respect to $\mu_d$



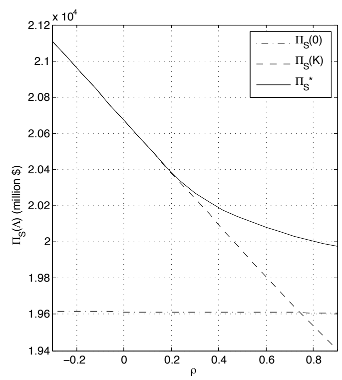

(b) Profit Comparison with Respect to $\mu_s$



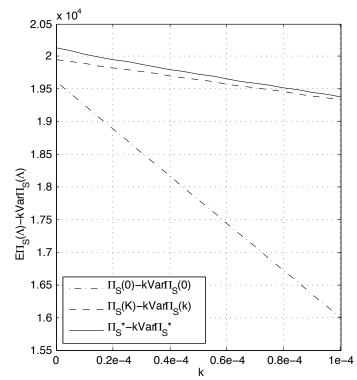

(c) Profit Comparison with Respect to $\sigma_d$




(d) Profit Comparison with Respect to $\sigma_s$




(e) Profit Comparison with Respect to $\rho$

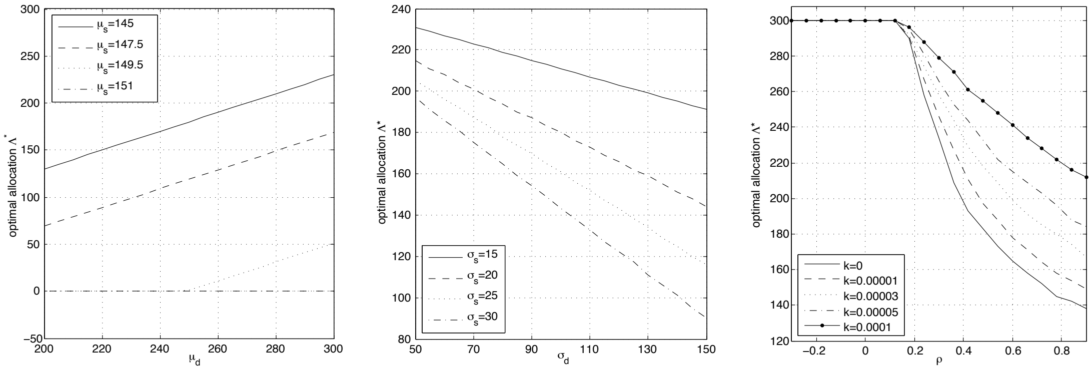



(f) Utility Comparison with Respect to $k$

Figure 4.2: Profit Comparision of $\Pi_S(0)$, $\Pi_S(K)$, and $\Pi_S(\Lambda^*)$ under an Open Spot Market

### The Impact of $\sigma_d$ and $\sigma_s$

Figure 4.2(c) and 4.2(d) demonstrate that: The supplier's optimal expected profit is decreasing in both the demand variability $\sigma_d$ and the spot price variability $\sigma_s$. The profit gap between the optimal allocation strategy and the total-contract strategy is increasing in both $\sigma_d$ and $\sigma_s$, with average profit improvements of 0.9% and 1.0%, respectively. The profit gap between the optimal allocation strategy and the total-spot strategy is decreasing in both $\sigma_d$ and $\sigma_s$, with average profit improvements of 2.70% and 2.74%, respectively. Figure 4.3(b) shows that under the assumed business environment, the commodity supplier should allocate less capacity to the contract channel as either the demand variability $\sigma_d$ or the spot price variability $\sigma_s$ increases,




(a) Optimal Allocation with Respect to $\mu_d$ and $\mu_s$

(b) Optimal Allocation with Respect to $\sigma_d$ and $\sigma_s$

(c) Optimal Allocation with Respect to $\rho$ and $k$

Figure 4.3: Comparative Statics for $\Lambda^*$ under an Open Spot Market

i.e., as the system becomes more volatile; this happens even when the average spot price $\mu_s$ is lower than the contract price $w$, as long as the supplier has a fixed risk aversion level $k$.

### The Impact of $\rho$ and $k$

Figure 4.2(e) and 4.2(f) demonstrate that: The supplier's optimal expected profit (utility) is decreasing in both the demand-price correlation $\rho$ and the risk aversion coefficient $k$. The profit gap between the optimal allocation strategy and the total-contract strategy is increasing in $\rho$ and decreasing in $k$, with average profit improvements of 1.76% and 0.49%, respectively. The profit gap between the optimal allocation strategy and the total-spot strategy is decreasing in $\rho$ and increasing in $k$, with average profit improvements of 3.62% and 11.20%, respectively. Figure 4.3(c) shows that the supplier should deploy his capacity fully to the contract channel if $\rho < 0$. When $\rho > 0$, though, the optimal contract channel allocation quantity $\Lambda^*$ is decreasing in $\rho$. In other words, the supplier should rely more on the spot channel as the contract demand and the spot price become more positively correlated. Meanwhile, we see that the supplier should allocate more capacity to the contract channel as he becomes more risk-averse, i.e., as $k$ increases.

# 4.5. Strategic Allocation under a Closed Spot Market

In Section 4.4, we discussed an open-spot-market setting in which the equilibrium spot price is exogenously determined. This is the case if there are sufficiently many other players in the spot market so that even the large supplier's participation will not impact the spot price formation. We believe this preliminary model is illuminative in general for firms facing the choice between a fixed price channel and a stochastic spot channel. In the actual iron ore industry, however, the spot price is usually the balanced outcome of the demand and supply dynamics in the market. For instance, Figure 4.4 below demonstrates the volume traded and the average spot price on China's iron ore spot market from 2001 to 2008; we can see a clear positive correlation between the quantity and the price. Hence, a supplier such as Rio Tinto usually has significant market power and its quantity decision is likely to affect the determination of the spot price. On one hand, if the supplier limits the quantity allocated to the contract channel, then the unsatisfied contract demand may switch to the spot market, driving up the spot price. On the other hand, if the supplier ships too much capacity to the spot market, then an increased supply level may instead drag down the equilibrium spot price. Therefore, the tradeoff has to be carefully decided upon on the basis of rigorous analytical modeling, which is indeed the focus of this section. We want to point out that the supplier's individual allocation behavior does not violate US antitrust law (Wolak 2001), nor does it conflict with any existing international regulation on the iron ore industry.

## 4.5.1 Endogenous Demand Curve and Exogenous Supply Curve

We first analyze a simple one-period setting in which the supplier's contract channel allocation decision affects the spot demand curve, but not the spot supply curve. At the beginning of the period, the supplier decides the maximum quantity $\Lambda$ that he would allocate to the contract channel. When demand $D$ is realized, it will first be satisfied through the contract channel $(\min(D, \Lambda))$; the leftover demand $(D - \Lambda)^+$,

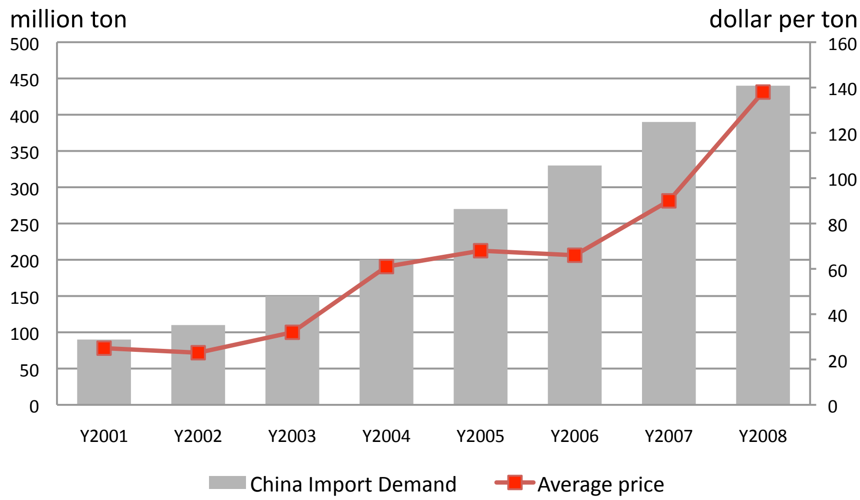



Figure 4.4: China Import Iron Ore Volume-Price Relationship

if any, will then shift to the spot market, and $\theta$ percent of that spot demand will be captured by the supplier again. The parameter $\theta$ here denotes the supplier's penetration power: if we assume the transaction is done unit by unit in the spot market, then with probability $\theta$ each unit demand will be satisfied by our supplier (see Wu *et al.* 2002 for a similar treatment). For this part only, we ignore the supplier's capacity limit, and assume there is sufficient leftover capacity to cover the switched-over demand $\theta(D-\Lambda)^+$. Under these assumptions, the equilibrium spot price $p_s$ is determined by the following equation:

$$p_s = a + b(D - \Lambda)^+ + \epsilon, \tag{4.5.1}$$

where $a, b > 0$ are constant coefficients, and $\epsilon$ represents some random noise. This formula reflects the fact that the equilibrium price is positively affected by the spot demand, given a fixed supply curve. Besides, it is also subject to some external randomness, which may potentially come from the supply side.

We justify Equation (4.5.1) using Figure 4.5. Note that the aggregate industry-wise demand $d$ is a step function of the spot price $p_s$: $d = (D-\Lambda)^+$ if $p_s$ is smaller than the end market price $p_m$, which is assumed to be constant; and $d = 0$ otherwise. We assume that there is a given market supply curve $p_s = f(S)$ where $f(\cdot)$ is increasing;

then, as the aggregated demand $d$ increases, the equilibrium spot price $p_s$ moves up along the supply curve. Therefore, if we assume a linear supply curve and $p_m$ is large enough, then the equilibrium price $p_s$ can be expressed by Equation (4.5.1). Note that we are modeling a situation where the spot demand is very price-inelastic. This reflects the practice that iron ore contributes about only 10% of the total cost of steel-making for large steel manufacturers (Jones 1986).

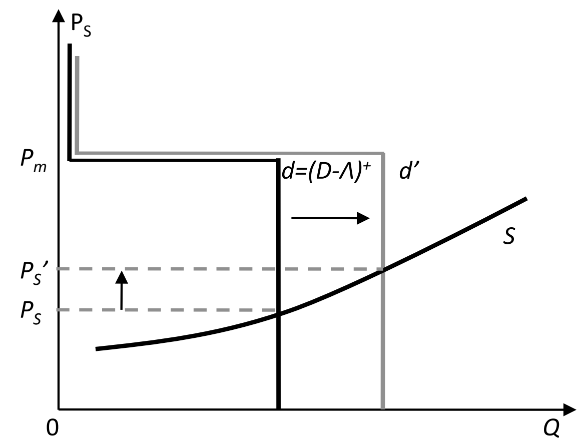



Figure 4.5: Determining of the Equilibrium Spot Price

Under the above setup, the supplier's optimization problem with a strategic allocation decision $\Lambda$ is denoted as

$$\pi_S = \max_{\Lambda \geq 0} \Pi_S(\Lambda), \tag{4.5.2}$$

where the objective function is

$$\Pi_S(\Lambda) = \mathbb{E}_{D,\epsilon}[(w-c)\min(D,\Lambda) + \theta(p_s - c)(D-\Lambda)^+], \tag{4.5.3}$$

and the spot price $p_s$ is given by Equation (4.5.1). Below, we first analyze some structural properties of $\Pi_S(\Lambda)$ assuming $D$ and $\epsilon$ are independent. Then, based on the structural results, we provide some policy recommendations to the supplier.

**Assumption 4.5.1.** *The demand distribution has a bounded support on* $[m, M]$*, where* $0 < m < M$*; and it also has an increasing failure rate (IFR), that is* $r(x) = \frac{\phi(x)}{1-\Phi(x)}$ *is increasing.*

It is well known that distributions such as the uniform, Erlang, normal, and truncated normal are all IFR (Porteus 2002). Hence this assumption will not significantly affect the applicability of our model.

**Lemma 4.5.2.** *IFR implies that* $r(x) = \frac{\phi(x)}{1-\Phi(x)} \leq \frac{1-\Phi(x)}{\int_x^M (1-\Phi(z))dz}$.

**Proposition 4.5.3.** *The supplier's objective function* $\Pi_S(\Lambda)$ *is quasiconvex in* $\Lambda$ *on interval* $[0, \infty]$.

**Proposition 4.5.4.** *Define* $\Xi = (\theta a + \bar{\theta} c - w)\mu_d + \theta b(\mu_d^2 + \sigma_d^2)$*, the supplier's optimal allocation policy is of an extreme type:*

*If* $\Xi \geq 0$*, then* $\Lambda^* = 0$ *and it is optimal to sell only through the spot channel;*

*If* $\Xi < 0$*, then* $\Lambda^* = \infty$ *and it is optimal to sell only through the contract channel.*

Proposition 4.5.4 says that the uncapacitated supplier facing an aggregated industry demand should either push all demand to the spot market by allocating no capacity to the contract channel or try to meet all demand through the contract channel, depending on the demand distribution ($\mu_d$, $\sigma_d$), the fixed contract price ($w$), the supplier's penetration power in the spot market ($\theta$), as well as the equilibrium spot price parameters ($a$, $b$). Despite risk, political, or reputational concerns, it is not optimal for an uncapacitated supplier with market power to mix sales between the contract channel and the spot channel, as long as his allocation decision does not impact the supply curve in the spot market.

**Corollary 4.5.5.** *Assuming* $a \geq c$*, the advantage of total withholding (represented by* $\Xi$*) is increasing in* $a$*,* $b$*,* $c$*,* $\theta$*,* $\sigma_d$*, and decreasing in* $w$*.*[4]

One interesting observation is that $\Xi$ is also increasing in $\sigma_d$, which means the larger the demand uncertainty is, the relatively more attractive total-spot policy (i.e., $\Lambda = 0$) becomes to the supplier. This echos our previous numerical observation for the open spot market case.

### Game Theory Model with Multiple Buyers

In the previous part, we replaced the buyers' ordering decision with an aggregated industry-wise demand and established an optimal extreme policy for the supplier. As

[4] $a \geq c$ implies the expected default spot price is higher than the supplier's unit cost.

a natural extension, we complicate our scenario here to some extent and look at a game theoretic setup in which the supplier makes an allocation decision, and multiple buyers determine their own contract quantities, if any. As demonstrated by Figure 4.6, the sequence of events within the period is as follows:

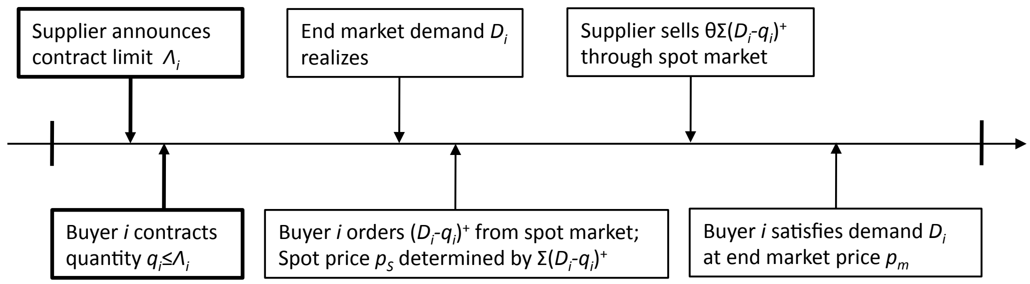



Figure 4.6: Sequence of Events for Strategic Withholding with Multiple Buyers

1. The supplier announces the contract quantity limit (allocation decision) $\Lambda_i$ for buyer $i$, $i = 1, 2, \cdots, N$.
2. Each buyer $i$ orders quantity $q_i(\leq \Lambda_i)$ through the contract channel.
3. End market demand $D_i$ realizes for each buyer $i$ according to distribution $\Phi_i(\cdot)$.
4. Buyer $i$ orders the shortage quantity $(D_i - q_i)^+$ from the spot market. The equilibrium spot price $p_s$ is determined by the formula.

$$p_s = a + b \sum_{i=1}^{N} (D_i - q_i)^+ + \epsilon. \tag{4.5.4}$$

5. Supplier sells $\theta \sum_{i=1}^{N} (D_i - q_i)^+$ through the spot market, where $\theta \in [0, 1]$ still denotes the supplier's market penetration power.
6. Buyer $i$ satisfies the end market demand $D_i$ at unit market price $p_m$.

Next, we analyze the model using backward induction – we first investigate the buyers' problem, then the supplier's.

### The Buyers' Problem

Buyer $i$ has only one nontrivial decision to make: contract quantity $q_i$ under the supplier's withholding decision $\Lambda_i$. He chooses the optimal $q_i \in [0, \Lambda_i]$ to maximize his expected profit, that is

$$\pi_{Bi}(\vec{\Lambda}) = \max_{0 \le q_i \le \Lambda_i} \Pi_{Bi}(q_i, \vec{q}_{-i}), \tag{4.5.5}$$

where the objective function is

$$\Pi_{Bi}(q_i, \vec{q}_{-i}) = \mathbb{E}_{\bar{D},\epsilon}\Big[p_m D_i - wq_i - (a + b\sum_{j=1}^{N}(D_j - q_j)^+ + \epsilon)(D_i - q_i)^+\Big]. \tag{4.5.6}$$

Given $\vec{q}_{-i}$, let $M_i$ be the maximum value that the demand $D_i$ can take; we then have the following result:

**Proposition 4.5.6.** $\Pi_{Bi}(q_i, \vec{q}_{-i})$ *is concave in* $q_i$ *and submodular in* $(q_i, q_j)$ *for all* $j \neq i$*. The unconstrained maximizer* $\hat{q}_i(\vec{q}_{-i})$ *satisfies the equation*

$$(a + b\sum_{j \neq i}\mathbb{E}(D_j - q_j)^+)(1 - \Phi_i(\hat{q}_i)) + 2b\int_{\hat{q}_i}^{M_i}(1 - \Phi_i(x))dx - w = 0. \tag{4.5.7}$$

Applying Topkis's Theorem by checking the sign of the cross-partials for different parameters, we can also obtain the following comparative statics result.

**Corollary 4.5.7.** *(Comparative Statics)* $\hat{q}_i$ *is decreasing in* $w$*; increasing in* $a$ *and* $b$*.*

We discuss the case of homogeneous buyers, in which all the $N$ buyers face i.i.d. demand $D$. It is then reasonable to assume that the supplier offers the same $\Lambda$ to all buyers. For this particular scenario, each buyer's contract policy can be characterized as follows:

**Proposition 4.5.8.** *When all the* $N$ *buyers face i.i.d. demand and the same contract limit* $\Lambda$ *from the supplier, their optimal order quantity* $q^*$ *is identical and given by* $q^* = \min(\hat{q}, \Lambda)$*, where* $\hat{q}$ *solves the following equation:*

$$(a + b(N-1)\mathbb{E}(D - \hat{q})^+)(1 - \Phi(\hat{q})) + 2b\int_{\hat{q}}^{M}(1 - \Phi(x))dx - w = 0. \tag{4.5.8}$$

### The Supplier's Problem

Now, at the beginning of the period, anticipating each buyer's optimal contract quantity $q_i^*(\Lambda_i)$, the supplier chooses the optimal $\vec{\Lambda}$ to maximize his expected profit, that is

$$\pi_S = \max_{\vec{\Lambda} \geq 0} \Pi_S(\vec{\Lambda}), \tag{4.5.9}$$

where the objective function

$$\Pi_S(\vec{\Lambda}) = \mathbb{E}_{\vec{D},\epsilon}\Big[(w-c)\sum_{i=1}^{N} q_i^*(\Lambda_i) + \theta(a + b\sum_{i=1}^{N}(D_i - q_i^*(\Lambda_i))^+ + \epsilon - c)\sum_{i=1}^{N}(D_i - q_i^*(\Lambda_i))^+\Big]$$

Similarly, we investigate the case with homogeneous buyers. From the previous discussion, we know that when all the buyers face i.i.d. demand, the supplier offers a unique contract limit $\Lambda$, and each buyer's optimal contract quantity is given by $q^* = \min(\hat{q}, \Lambda)$. The supplier achieves identical expected profit by offering $\Lambda \in [\hat{q}, \infty)$ (since buyers would always contract $\hat{q}$). Therefore, we only need to investigate $\Lambda \in [0, \hat{q}]$, in which case each buyer contracts exactly $\Lambda$. Based on this analysis, the supplier is actually solving the following simplified optimization:

$$\hat{\pi}_S = \max_{0 \leq \Lambda \leq \hat{q}} \hat{\Pi}_S(\Lambda), \tag{4.5.10}$$

where the new objective function

$$\hat{\Pi}_S(\Lambda) = \mathbb{E}_{\vec{D},\epsilon}[(w-c)N\Lambda + \theta(a + b\sum_{i=1}^{N}(D_i - \Lambda)^+ + \epsilon - c)\sum_{i=1}^{N}(D_i - \Lambda)^+], \tag{4.5.11}$$

and all the $D_i$'s are i.i.d.

**Proposition 4.5.9.** *Assuming $a \geq c$, $\hat{\Pi}_S(\Lambda)$ is convex in $\Lambda$, which implies the original $\Pi_S(\Lambda)$ is quasiconvex on $[0, \infty)$.*

**Proposition 4.5.10.** *Let $\Xi' = N\theta(Nb\mu_d^2 + (a-c)\mu_d + b\sigma_d^2) - \hat{\Pi}_S(\hat{q})$; the supplier's optimal allocation policy is of an extreme type:*

*If $\Xi' \geq 0$, then $\Lambda^* = 0$ and total-spot is the optimal strategy; there exists a*

$\Lambda' \in [0, \hat{q}]$ *such that allocating with any* $\Lambda \in [0, \Lambda']$ *leads to a higher expected profit for the supplier than choosing total-contract.*

*If* $\Xi' < 0$*, then* $\Lambda^* = \infty$ *and total-contract is the optimal strategy.*

Proposition 4.5.10 demonstrates that even with a complicated game theoretic setting, a supplier facing an endogenous spot demand curve and an exogenous spot supply curve would still follow a bang-bang allocation policy – either choose a total-contract strategy or a total-spot strategy. However, in contrast to the aggregated-demand case in which the supplier sells through only one channel, here as long as there are buyers facing a realized demand larger than the quantity they previously contracted, the supplier may still participate in the spot market even if he allocated all the capacity to the contract channel first.

### 4.5.2 Endogenous Demand and Supply Curves

In Section 4.5.1, we established the optimality of an extreme policy for the supplier assuming that the supplier's allocation decision affects only the industry demand curve in the spot market, not the supply curve. Under some circumstances, due to the restriction of internal stockpiling space and the nonzero delivery leadtime, the supplier may need to ship the quantity to the local spot market before the spot trading takes place, which will effectively shift the spot supply curve as well. Hence, in this section we relax the aforementioned limiting assumptions and tackle the most general yet most complicated dual-channel commodity selling problem that a capacitated supplier faces. Again, we investigate a single period problem with aggregated contract-channel demand $D$, which is stochastic. The supplier with a capacity limit $K$ makes two sequential decisions at two stages within the period.

Stage One: The supplier allocates capacity $\Lambda$ ($\leq K$) to the contract channel before knowing the final demand. In practice, this could mean setting an upper-bound for the total contract volume to be executed. As with the previous version of the model, we assume that the contract channel has a higher priority to the downstream industry than the spot channel; that is, the demand is first satisfied through the forward contract, and then $\theta$ percent of the leftover demand $(D - \Lambda)^+$ goes to the spot market later on. Here the constant $\theta$ ($\leq 1$) suggests that a fixed portion $(1 - \theta)$ of the leftover demand would be absorbed at some place other than the spot market, and its value can be estimated from historical data. We also assume that the contract channel quantity incurs a unit holding cost of $h$ since it needs to be stockpiled and ready before the orders arrive.

Stage Two: The supplier ships quantity $\Lambda_S$ ($\leq K - \min(D, \Lambda)$) from his leftover capacity to sell in the spot market once the contract channel demand $D$ is realized.

Given that the supplier has decided on the two quantity decisions $\Lambda$ and $\Lambda_S$, we now explain how the demand and supply curves in the spot market would be determined (Figure 4.7).

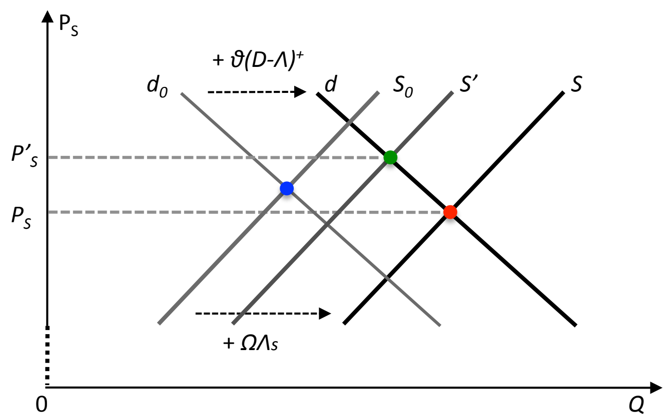



Figure 4.7: Determining of the Equilibrium Price

**The Demand Curve** Different from the setting described in Section 4.5.1, here we assume that there is a default spot demand curve $d_0 : Q = \gamma - \delta p_s$ ($\gamma, \delta > 0$) before the leftover demand $\theta(D - \Lambda)^+$ from the contract channel arrives at the spot market. We add this complexity to better reflect the industry practice that there are usually some external spot buyers other than the large buyers from the contract channel. As Graeme Stanway, chief iron ore consultant at Rio Tinto, put it, "The long-term contract is more geared to the larger (steel) mills particularly Japanese, Korean, and large Chinese players such as Bao Steel.$\cdots$The spot market is a mechanism that allows a broader range of steel mills to access high quality iron ore (G. Stanway, personal communication, November 24, 2010)."

Given $d_0$, which is price sensitive, and the switched-over demand $\Delta d = \theta(D - \Lambda)^+$ from large buyers, which is assumed to be price inelastic, the final demand curve in the spot market is given by

$$d: \quad Q = \gamma + \theta(D - \Lambda)^+ - \delta p_s. \tag{4.5.12}$$

**The Supply Curve** Similarly to the demand side, we assume that before our supplier participates in the spot market, there is a default industry supply curve given by $S_0 : Q = -\alpha + \beta p_s$ ($\alpha, \beta > 0$). After the amount $\Lambda_S$ is added to the market, the supply curve will be shifted to the right by $\Omega\Lambda_S$ ($\Omega > 0$) and become

$$S: \quad Q = -\alpha + \Omega\Lambda_S + \beta p_s. \tag{4.5.13}$$

The parameter $\Omega$ indirectly reflects the potential reaction from other spot suppliers to our large supplier's quantity decision. In particular, $\Omega \leq 1$ suggests that the quantity provided by other suppliers will decrease, possibly due to resource competition. Instead, $\Omega > 1$ implies that the large supplier's participation in the spot market may lead to a market-following effect among the small spot suppliers.

**The Equilibrium Price** Given the above discussion, the equilibrium spot price $p_s$ can be determined by equating the demand $d$ and the supply $S$. Specifically, solving Equation (4.5.12) and (4.5.13) jointly, we obtain the following:

$$p_s = \frac{\alpha + \gamma + \theta(D - \Lambda)^+ - \Omega\Lambda_S}{\beta + \delta}. \tag{4.5.14}$$

For the ease of notation, we let $a = \frac{\alpha+\gamma}{\beta+\delta}$, $b = \frac{\theta}{\beta+\delta}$, and $g = \frac{\Omega}{\beta+\delta}$ ($a, b, g > 0$); then, the above formula can be further simplified to

$$p_s = a + b(D - \Lambda)^+ - g\Lambda_S. \tag{4.5.15}$$

Note that the random term $\epsilon$ in Formula 4.5.1 is now replaced by $-g\Lambda_S$, which reflects the impact of the supplier's quantity decision on the spot supply curve. One underlying assumption here is that the large supplier has the lowest unit production cost $c$ (due to economies of scale) among all the spot suppliers, and that $c$ is even lower than the default spot price $p_s^0$. Hence, the supplier's spot quantity $\Lambda_S$ is inelastic in price and can be represented by a vertical line in the price-quantity quadrant.

Similarly, the switched-over demand $\theta(D-\Lambda)^+$ from the large contract buyers is also inelastic to the spot price and can be depicted by a vertical line. This assumption explains the parallel shift of the default spot demand and supply curves.

In the following section, we analyze the supplier's decision-making process using the standard method of backward induction.

**Stage 2: The Determination of $\Lambda_S$**

Given the capacity limit $K$, the first stage allocation decision $\Lambda$, and the actual realized industry-wise demand $D$, the supplier chooses a spot quantity $\Lambda_S$ to maximize his second stage expected profit:

$$\pi_S^2(\Lambda, D) = \max_{0 \le \Lambda_S \le K - \min(D,\Lambda)} \Pi_S^2(\Lambda, D, \Lambda_S), \tag{4.5.16}$$

where the objective function is given by

$$\Pi_S^2(\Lambda, D, \Lambda_S) = (p_s - c)\Lambda_S = [a + b(D-\Lambda)^+ - g\Lambda_S - c]\Lambda_S. \tag{4.5.17}$$

**Proposition 4.5.11.** *$\Pi_S^2(\Lambda, D, \Lambda_S)$ is concave in $\Lambda_S$ with the global maximizer $\hat{\Lambda}_S$ and the global maximum $\hat{\pi}_S^2$ given by $\hat{\Lambda}_S = \frac{1}{2g}(a+b(D-\Lambda)^+-c)$, $\hat{\pi}_S^2 = \frac{1}{4g}(a+b(D-\Lambda)^+-c)^2$.*

**Proposition 4.5.12.** *Assuming $0 \le \frac{a-c}{2g} \le K$, based on the value of $\Lambda$ and $D$, and that $0 \le \Lambda_S \le K - \min(D,\Lambda)$, the supplier's optimal decision of the Stage 2 problem is characterized by Table 4.2.*

Proposition 4.5.11 and 4.5.12 establish that the supplier's second stage problem is well behaved, and that the optimal spot participation quantity $\Lambda_S^*$ is determined according to a modified-base-stock type of policy. Intuitively, we can see that the unconstrained optimal spot quantity $\hat{\Lambda}_S$ is increasing in $D$, $a$, and $b$, and decreasing in $\Lambda$, $c$, and $g$.

**Stage 1: The Determination of $\Lambda$**

At Stage 1, with the full contingent plan for the second stage based on the analysis in Section 4.5.2, the supplier chooses $\Lambda \in [0, K]$ to maximize his expected profit over

Table 4.2: Optimal Solution of the Stage Two Problem

| Case | Range of $\Lambda$ | Range of $D$ | Optimal Decision $\Lambda_S^*$ |
|---|---|---|---|
| $(i)$ | | $(0, \Lambda]$ | $\frac{1}{2g}(a-c)$ |
| $(ii)$ | $[0,\ K-\frac{a-c}{2g}]$ | $(\Lambda,\ \frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))]$ | $\frac{1}{2g}(a+b(D-\Lambda)-c)$ |
| $(iii)$ | | $(\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c)),\ M)$ | $K-\Lambda$ |
| $(iv)$ | | $(0,\ K-\frac{a-c}{2g}]$ | $\frac{1}{2g}(a-c)$ |
| $(v)$ | $(K-\frac{a-c}{2g},\ K]$ | $(K-\frac{a-c}{2g},\ \Lambda]$ | $K-D$ |
| $(vi)$ | | $(\Lambda,\ M)$ | $K-\Lambda$ |

the entire period:

$$\pi_S^1 = \max_{0\le\Lambda\le K} \Pi_S^1(\Lambda), \tag{4.5.18}$$

where the objective function is given by

$$\begin{aligned} \Pi_S^1(\Lambda) &= \mathbb{E}_D\ \{(w-c)\min(D,\Lambda)+\pi_S^2(\Lambda,D)\} \\ &= \mathbb{E}_D\ \{(w-c)\min(D,\Lambda)+\Pi_S^2(\Lambda,D,\Lambda_S^*(\Lambda,D))\}, \end{aligned} \tag{4.5.19}$$

and $\Lambda_S^*$ is described in Proposition 4.5.12. Depending on the specific range of $\Lambda$ and $D$, we can further expand the optimization problem as follows:

$$\pi_S^1 = \max\{\max_{0\le\Lambda\le K-\frac{a-c}{2g}} \Pi_{S,1}^1(\Lambda),\ \max_{K-\frac{a-c}{2g}<\Lambda\le K} \Pi_{S,2}^1(\Lambda)\}, \tag{4.5.20}$$

where

$$\begin{aligned} \Pi_{S,1}^1(\Lambda) &= \mathbb{E}_D\ (w-c)\min(D,\Lambda)-h\Lambda \\ &+\int_0^{\Lambda}\frac{(a-c)^2}{4g}\phi(x)dx+\int_{\Lambda}^{\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))}\frac{(a+b(x-\Lambda)-c)^2}{4g}\phi(x)dx \\ &+\int_{\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))}^{M}(K-\Lambda)[a+b(x-\Lambda)-g(K-\Lambda)-c]\phi(x)dx, \end{aligned} \tag{4.5.21}$$

$$\begin{aligned}\Pi^1_{S,2}(\Lambda) &= \mathbb{E}_D\ (w-c)\min(D,\Lambda) - h\Lambda \\ &+ \int_0^{K-\frac{a-c}{2g}} \frac{(a-c)^2}{4g}\phi(x)dx + \int_{K-\frac{a-c}{2g}}^{\Lambda} (K-x)[a-g(K-x)-c]\phi(x)dx \\ &+ \int_{\Lambda}^{M} (K-\Lambda)[a+b(x-\Lambda)-g(K-\Lambda)-c]\phi(x)dx. \end{aligned} \tag{4.5.22}$$

**Proposition 4.5.13.** *Assuming $b \le g$ and $a \le w$, there exist mild conditions on the demand distribution[5] such that the supplier's expected profit function $\Pi^1_S(\Lambda)$ is convex-concave in $\Lambda$ on interval $[0, K]$ with at most one interior maximizer $\hat{\Lambda}$, which would be the unique solution to either $\Pi^1_{S,1}(\Lambda)' = 0,\ (\Lambda \in [0, K-\frac{a-c}{2g}])$ or $\Pi^1_{S,2}(\Lambda)' = 0,\ (\Lambda \in (K-\frac{a-c}{2g}, K])$.*

Proposition 4.5.13 establishes a very important structural property of the supplier's first stage profit function. Under convex-concavity, we need to compare at most three allocation decisions: the two extreme values $\Lambda = 0$ and $\Lambda = K$, as well as the interior local maximizer $\hat{\Lambda}$ (if it exists). Whichever leads to the highest expected profit should be chosen as the optimal allocation decision. This strategy is expressed formally in the following corollary:

**Corollary 4.5.14.** *The supplier's optimal first stage allocation decision $\Lambda^*$ is characterized by*

$$\Lambda^* = \begin{cases} \hat{\Lambda}, & \text{if } \hat{\Lambda} \text{ exists and } \Pi^1_S(\hat{\Lambda}) \ge \Pi^1_S(0) \\ 0, & \text{if } \Pi^1_S(0) > \Pi^1_S(\hat{\Lambda}) \text{ and } \Pi^1_S(0) \ge \Pi^1_S(K) \\ K, & \text{if } \hat{\Lambda} \text{ does not exist and } \Pi^1_S(K) > \Pi^1_S(0) \end{cases} \tag{4.5.23}$$

*where*

$$\Pi^1_S(0) = \int_0^{\frac{(2gK-(a-c))}{b}} \frac{(a+bx-c)^2}{4g}\phi(x)dx + \int_{\frac{(2gK-(a-c))}{b}}^{M} K(a+bx-gK-c)\phi(x)dx, \tag{4.5.24}$$

$$\Pi^1_S(K) = (w-c)[\int_0^K x\phi(x)dx + K(1-\Phi(K))] - hK + \frac{(a-c)^2}{4g}\Phi(K-\frac{a-c}{2g})$$

[5]Refer to Assumption B.0.3 in the proof.

$$+\int_{K-\frac{a-c}{2g}}^{K}(K-x)[a-g(K-x)-c]\phi(x)dx. \quad (4.5.25)$$

A further inspection of the validity of Proposition 4.5.13 (proof available in appendix) reveals that the concavity part of the result relies heavily on the condition that $b \leq g$, i.e., $\theta \leq \Omega$. In other words, if the shifting effect of the supplier's second stage quantity decision $\Lambda_S$ on the default spot supply curve is stronger than the shifting effect of the unfulfilled first stage demand $(D-\Lambda)^+$ on the default spot demand curve, then it is likely that the second half of the expected profit function is concave and that it is optimal for the supplier to balance the sales between the contract channel and the spot channel, i.e., $\Lambda^* \in (0, K)$. If $b > g$ ($\theta > \Omega$), however, it is more likely that the supplier would face a convex profit function, in which case an extreme policy similar to the one in the previous section would be optimal. In the actual iron ore industry, since buyers usually have alternatives of sourcing from medium-size ore suppliers in India or Africa as well, $\theta$ is certainly less than 1, and thus we expect $\theta \leq \Omega$ to hold.

**Proposition 4.5.15.** *(Comparative Statics) The potential interior maximizer $\hat{\Lambda}$ and the three function values, $\Pi_S^1(\hat{\Lambda})$, $\Pi_S^1(0)$, and $\Pi_S^1(K)$, observe the monotonicity property described in Table 4.3:*

Table 4.3: Comparative Statics (↑: increasing, ↓: decreasing, –: irrelevant)

| Parameters | $\hat{\Lambda}$ | $\Pi_S^1(\hat{\Lambda})$ | $\Pi_S^1(0)$ | $\Pi_S^1(K)$ |
|---|---|---|---|---|
| $a$ (i.e., $\alpha$,$\gamma$) | ↓ | ↑ | ↑ | – |
| $b$ (i.e., $\theta$) | ↓ | ↑ | ↑ | – |
| $g$ (i.e., $\Omega$) | ↑ | ↓ | ↓ | – |
| $w$ | ↑ | ↑ | – | ↑ |
| $c$ | ↓ | ↓ | ↓ | ↓ |
| $h$ | ↓ | ↓ | – | ↓ |

The monotonicity with respect to $\mu_d$, $\sigma_d$, $\beta$ and $\delta$ (demand/supply curve price elasticity coefficients) is not analytically definable and will be discussed through numerical analysis in Section 4.5.4.

### 4.5.3 Special Cases

In the previous section, we formulated the supplier's capacity allocation problem for the most general case, where the contract channel demand $D$ is a continuous random variable with a given distribution. We also derived some analytical results showing the structure of the supplier's optimal policy. However, due to the complexity of the general model, we were not able to demonstrate the strategy in a closed-form manner. In this section, we concentrate on two special cases, in which the aggregate contract demand $D$ is either deterministic or two-point distributed. This allows us to discuss the corresponding optimal policies and managerial insights in a more specific and illustrative way.

#### Deterministic Demand

We first investigate the simplest scenario in which $D$ is deterministic and therefore the entire system contains no randomness. To reduce the number of unnecessary contingencies, we assume $D \leq K$; the previous condition that $\frac{a-c}{2g} \leq K$ also applies. Obviously, the first stage contract allocation quantity $\Lambda$ will not exceed $D$. The following proposition delineates the detailed contingency map for the optimal decisions in the two stages. We can see that even under this deterministic case, the first stage allocation quantity may not be extreme.

**Proposition 4.5.16.** *Define* $\hat{\Lambda} = \frac{1}{2(g-b)}[w - a - bD - h + (2g-b)K]$, *we have* $\frac{2gK-(a+bD-c)}{2g-b} \leq \hat{\Lambda} \leq D$ *and the supplier's optimal allocation policy under a deterministic demand $D$ is given by:*

*If* $D \leq \min(\frac{4g(w-c-h)-2b(a-c)}{b^2}, K - \frac{a-c}{2g})$, *then* $\Lambda^* = D,\ \Lambda_S^* = \frac{a-c}{2g}$;

*If* $\frac{4g(w-c-h)-2b(a-c)}{b^2} < D \leq K - \frac{a-c}{2g}$, *then* $\Lambda^* = 0,\ \Lambda_S^* = \frac{a+bD-c}{2g}$;

*If* $\max(K - \frac{a-c}{2g}, K - \frac{a+h-w}{2g-b}) \leq D \leq K - \frac{a-c}{b} + \frac{(2g-b)(w-c-h)}{b^2}$, *we have:*

*If* $(w-c-h)\hat{\Lambda} + (a + b(D-\hat{\Lambda}) - g(K-\hat{\Lambda}) - c)(K-\hat{\Lambda}) \geq \frac{(a+bD-c)^2}{4g}$, *then* $\Lambda^* = \hat{\Lambda},\ \Lambda_S^* = K - \hat{\Lambda}$;

*If else, then* $\Lambda^* = 0,\ \Lambda_S^* = \frac{a+bD-c}{2g}$;

*Otherwise, we have:*

*If* $(w-c-h)D+(a-g(K-D)-c)(K-D) \geq \frac{(a+bD-c)^2}{4g}$, *then* $\Lambda^* = D,\ \Lambda_S^* = K-D$;

*If else, then* $\Lambda^* = 0,\ \Lambda_S^* = \frac{a+bD-c}{2g}$.

We can see from the third scenario above that even when the contract demand is deterministic, the supplier's first stage allocation decision may not be totally extreme: there are circumstances under which the supplier should satisfy only part of the demand in the contract channel and save the rest of the capacity for the spot market. Hence, when the equilibrium spot price is completely endogenous, demand uncertainty is not the only reason for the supplier to adopt a dual-channel strategy, though the uncertainty may have an impact on the specific allocation quantities (will be shown in numerical analysis).

**Two-Point Demand**

We next discuss another commonly adopted setting in which the demand is two-point distributed, representing both an optimistic scenario and a pessimistic scenario. In particular, we assume that $D = D_H$ with probability $p$, and that $D = D_L$ with probability $\bar{p} = 1 - p$, where $0 < D_L \leq K - \frac{a-c}{2g} < D_H \leq K$ and $0 < p < 1$. In this stochastic case, the supplier's second stage allocation decision $\Lambda_S$ is still given by Proposition 4.5.12. We know with certainty that the supplier's first stage decision $\Lambda \in [0, D_H]$; however, since Assumption B.0.3 (see appendix) that supports Proposition 4.5.13 in the previous section is no longer valid here, the structure of the optimal policy will consequently be different. More explicitly, the supplier's optimal allocation strategy is given by Proposition 4.5.17 below:

**Proposition 4.5.17.** *Assume* $b \leq g$, *let* $\gamma_1 = \frac{2gK-(a-c)-bD_H}{2g-b}$, $\gamma_2 = \frac{2gK-(a-c)-bD_L}{2g-b}$, $\hat{\Lambda}_1 = \frac{p(w-a-bD_H+(2g-b)K)-h}{2p(g-b)}$, $\hat{\Lambda}_2 = \frac{2g[(w-c-h)-p(a+bD_H-c-(2g-b)K)]-\bar{p}b(a+bD_L-c)}{4pg(g-b)-\bar{p}b^2}$, *and* $\mu_d = pD_H + \bar{p}D_L$, *then* $\gamma_1 < D_H$, $\gamma_2 \geq D_L$, $\gamma_1 < K - \frac{a-c}{2g} \leq \gamma_2$, *and the supplier's first stage allocation decision* $\Lambda$ *is given by:*

*(i) When* $\gamma_1 \leq D_L$, *if* $p \in [0, (\frac{b}{2g-b})^2]$, *then* $\Pi_S^1(\Lambda)$ *is convex on* $[0, D_L]$ *and concave on* $[D_L, D_H]$*:*

*if* $\hat{\Lambda}_1 \in [D_L, D_H]$, *then* $\Lambda^* = \hat{\Lambda}_1$ *if* $\Pi_S^1(\hat{\Lambda}_1) \geq \Pi_S^1(0)$, *and* $\Lambda^* = 0$ *otherwise;*

*if* $\hat{\Lambda}_1 \notin [D_L, D_H]$, *then* $\Lambda^* = 0, D_L$, *or* $D_H$, *whichever leads to the highest profit.*

*If* $p \in ((\frac{b}{2g-b})^2, 1]$, *then* $\Pi_S^1(\Lambda)$ *is convex on* $[0, \gamma_1]$, *concave on* $[\gamma_1, D_L]$ *and* $[D_L, D_H]$*:*

*if* $\hat{\Lambda}_1 \in [D_L, D_H]$, *then* $\Lambda^* = \hat{\Lambda}_1$ *if* $\Pi_S^1(\hat{\Lambda}_1) \geq \Pi_S^1(0)$, *and* $\Lambda^* = 0$ *otherwise;*

*if* $\hat{\Lambda}_2 \in [\gamma_1, D_L]$, *then* $\Lambda^* = \hat{\Lambda}_2$ *if* $\Pi_S^1(\hat{\Lambda}_2) \geq \Pi_S^1(0)$, *and* $\Lambda^* = 0$ *otherwise;*

*else,* $\Lambda^* = 0, D_L$, *or* $D_H$, *whichever leads to the highest profit.* $\Pi_S^1(\Lambda)$ *is given by*

*Equation (B.0.43) in the appendix.*

*(ii) When $\gamma_1 > D_L$, $\Pi_S^1(\Lambda)$ is convex on $[0, D_L]$, $[D_L, \gamma_1]$, and concave on $[\gamma_1, D_H]$:*

*if $\hat{\Lambda}_1 \in [\gamma_1, D_H]$, then $\Lambda^* = \hat{\Lambda}_1, 0$, or $D_L$, whichever leads to the highest profit;*

*if $\hat{\Lambda}_1 \notin [\gamma_1, D_H]$, then $\Lambda^* = 0, D_L$, or $D_H$, whichever leads to the highest profit.*

*$\Pi_S^1(\Lambda)$ is given by Equation (B.0.44) in the appendix.*

Different from the case where the contract channel demand is continuously distributed and only one interior local maximum may exist, here the lower demand value $D_L$ can be another local maximal point that the supplier should evaluate.

### 4.5.4 Numerical Analysis

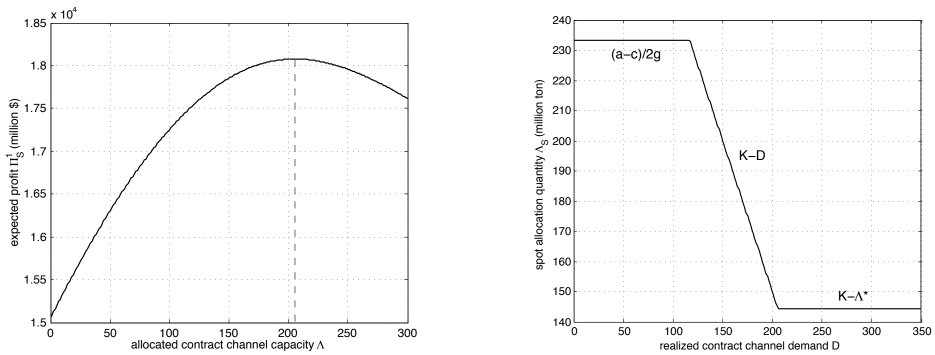
In the previous section we provided an analytical discussion of the commodity supplier's capacity allocation strategy during the two decision stages. We also provided some comparative statics results in terms of how the optimal allocation quantity should respond to the change of several parameter values. In this part, we will generate additional managerial insights by conducting a thorough numerical analysis. We focus on the most comprehensive case where the supplier's quantity decision can affect both the spot demand curve and the spot supply curve, which consequently determine the equilibrium spot price.

Table 4.4: Benchmark Values of Model Parameters (units: $K, \alpha, \gamma, \mu_d, \sigma_d$: million ton; $w, c, h$: dollar per ton; $\beta, \delta$: million ton per dollar; $\theta, \Omega$: no unit)

| parameter | value | parameter | value | parameter | value |
|---|---|---|---|---|---|
| $K$ | 300 | $w$ | 150 | $c$ | 80 |
| $h$ | 6 | $\alpha$ | 20 | $\beta$ | 3 |
| $\gamma$ | 1180 | $\delta$ | 5 | $\Omega$ | 1 |
| $\theta$ | 0.8 | $\mu_d$ | 250 | $\sigma_d$ | 100 |

Table 4.4 above shows the benchmark parameter values, which are partially based on industry practice and partially based on reasonable estimates. Figure 4.8(a) plots the supplier's expected profit versus his first stage allocation decision $\Lambda$, and we can see it is optimal for the supplier to allocate $\Lambda^* = 205.7$ million tons to the

contract channel in the first stage before seeing the demand. Figure 4.8(b) then demonstrates the optimal quantity to be shipped to the spot market during the second stage contingent upon the realized contract channel demand $D$.



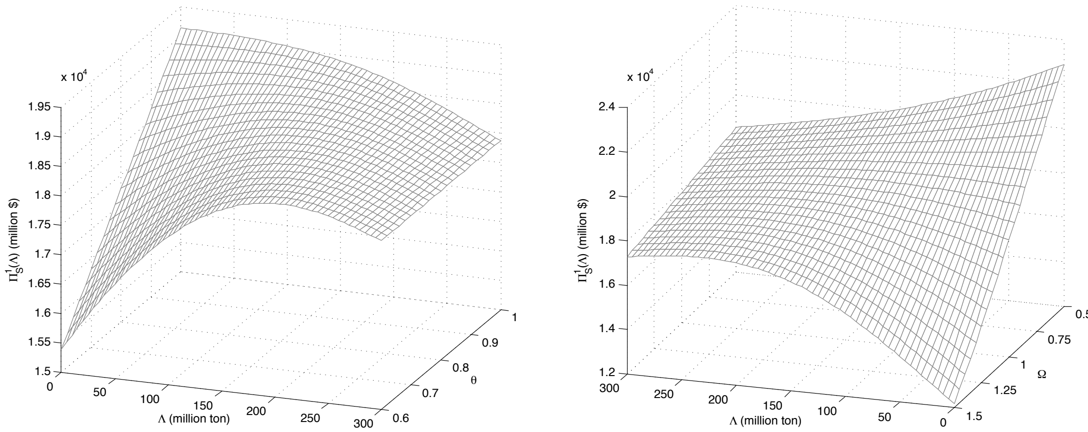

(a) Expected profit with respect to $\Lambda$ (b) Optimal $\Lambda_S$ with respect to $D$

Figure 4.8: The Supplier's Optimal Allocation Decisions for the Benchmark Case

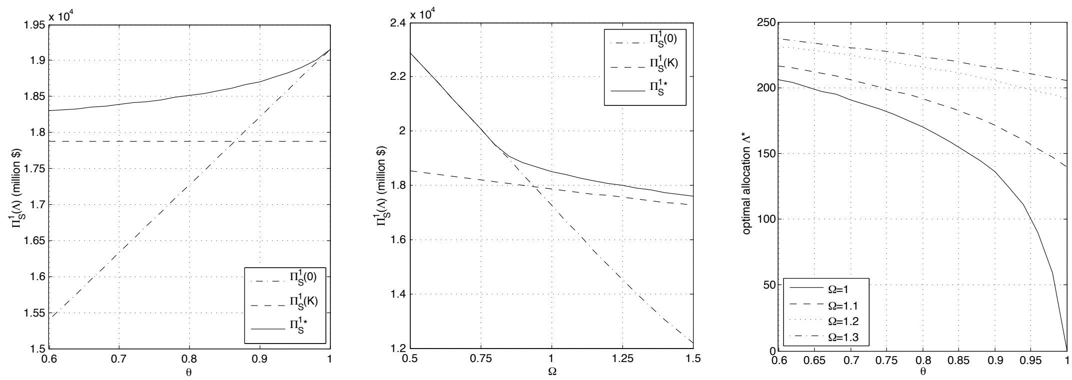
Next, we investigate the sensitivity of the supplier's first stage allocation decision, as well as his total expected profit with respect to several key model parameters.

### The Impact of $\theta$ and $\Omega$

Figure 4.9 shows the impact of the demand switch ratio $\theta$ and the spot supply emulation factor $\Omega$ on the supplier's expected profit and his optimal first stage allocation decision. Let us refer to the policy of allocating all capacity to the contract channel in the first stage (i.e., $\Lambda = K$) as the *total-contract* policy, and the one of reserving all capacity for the spot market (i.e., $\Lambda = 0$) as the *total-spot* policy. Panels (a) and (b) provide a 3-D overview of the results. Panels (c) and (d) demonstrate that: The optimal expected profit is increasing in $\theta$ and decreasing in $\Omega$. The profit gap between the optimal allocation strategy (with a mixed portfolio) and the total-contract strategy is increasing in $\theta$ and decreasing in $\Omega$, with average profit improvements of 3.94% and 7.23%, respectively. The profit gap between the optimal allocation strategy and the total-spot strategy is decreasing in $\theta$ and increasing in $\Omega$, with average profit improvements of 7.95% and 13.27%, respectively. Finally, Panel (e) shows that the optimal contract channel allocation quantity $\Lambda^*$ is decreasing in $\theta$ and increasing in $\Omega$, which is consistent with the result in Proposition 4.5.15.




(a) Expected profit with respect to $\Lambda$ and $\theta$ (b) Expected profit with respect to $\Lambda$ and $\Omega$

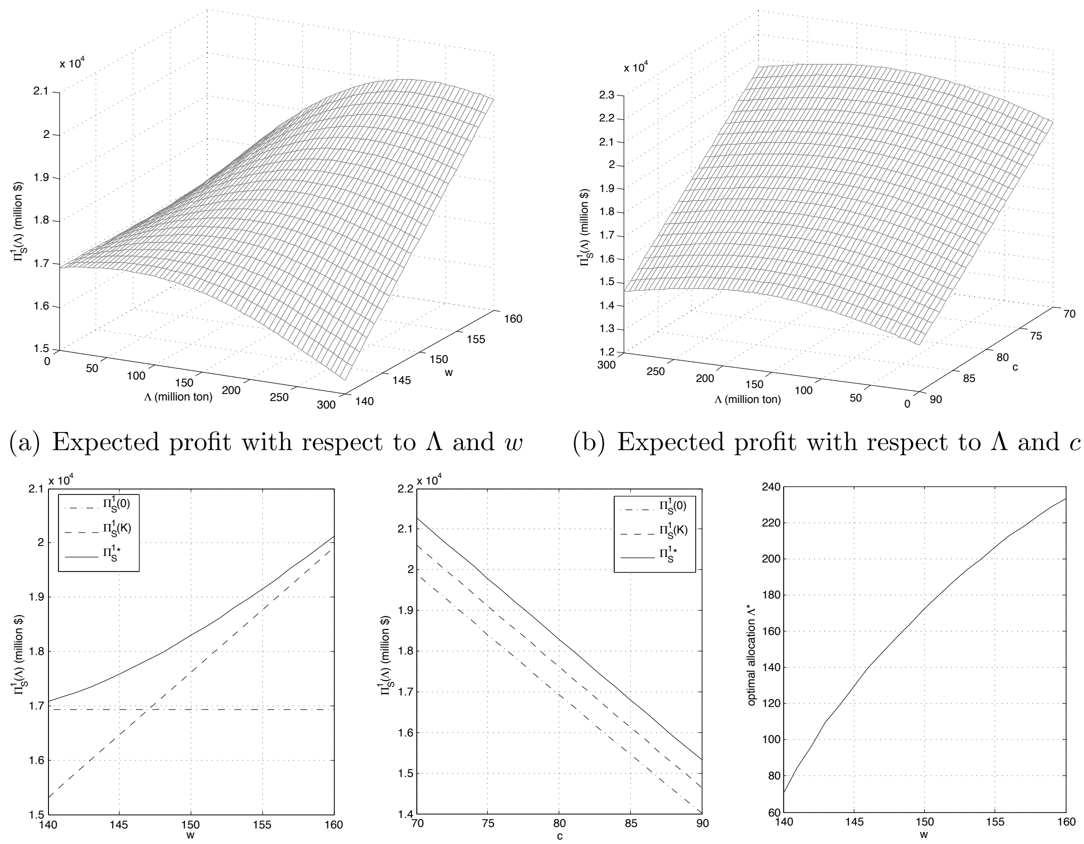



(c) Profit Comparison with Respect to $\theta$ (d) Profit Comparison with Respect to $\Omega$ (e) Optimal Allocation with Respect to $\theta$ and $\Omega$

Figure 4.9: Impact of $\theta$ and $\Omega$ on Expected Profit and Optimal Allocation Decision

### The Impact of $w$ and $c$

Figure 4.11 shows the impact of the contract channel price $w$ and the unit production cost $c$ on the supplier's expected profit and his optimal first stage allocation decision. Panels (a) and (b) provide a 3-D overview. Panels (c) and (d) demonstrate that: The expected total profit is increasing in $w$ and decreasing in $c$, which is quite intuitive. The profit gap between the optimal allocation strategy and the total-contract strategy is decreasing in $w$ and constant in $c$, with average profit improvements of 4.71% and 3.88%, respectively. The profit gap between the optimal allocation strategy and the total-spot strategy is increasing in $w$ and constant in $c$, with average profit




(a) Expected profit with respect to $\Lambda$ and $w$ (b) Expected profit with respect to $\Lambda$ and $c$

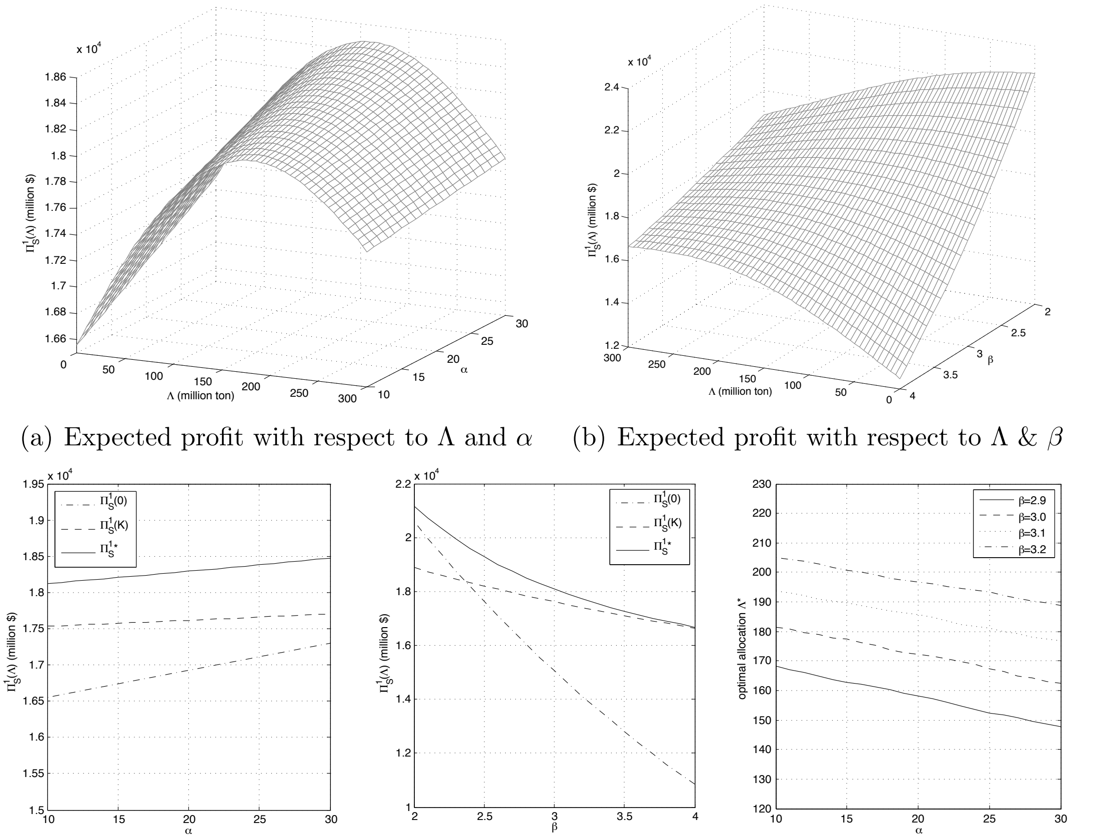
(c) Profit Comparison with Respect to $w$ (d) Profit Comparison with Respect to $c$ (e) Optimal Allocation with Respect to $w$ and $c$

Figure 4.10: Impact of $w$ and $c$ on Expected Profit and Optimal Allocation Decision

improvements of 8.73% and 8.09%, respectively. Panel (e) further shows that the optimal contract allocation quantity $\Lambda^*$ is increasing in $w$ and independent of $c$.

### The Impact of $\alpha$ ($\gamma$) and $\beta$ ($\delta$)

From the equilibrium spot price Equation 4.5.14, we can see that $\alpha$ and $\gamma$ play equivalent roles in the system, as do the quantity-price sensitivity coefficients $\beta$ and $\delta$. Hence we only select one parameter from each pair, $\alpha$ and $\beta$ in particular, to investigate their impacts on the expected profit as well as the optimal allocation decision. Panels (a) and (b) in Figure 4.11 summarize the results in a 3-D form. Panels (c) and (d) demonstrate that: The optimal expected profit is increasing in




(a) Expected profit with respect to $\Lambda$ and $\alpha$

(b) Expected profit with respect to $\Lambda$ & $\beta$

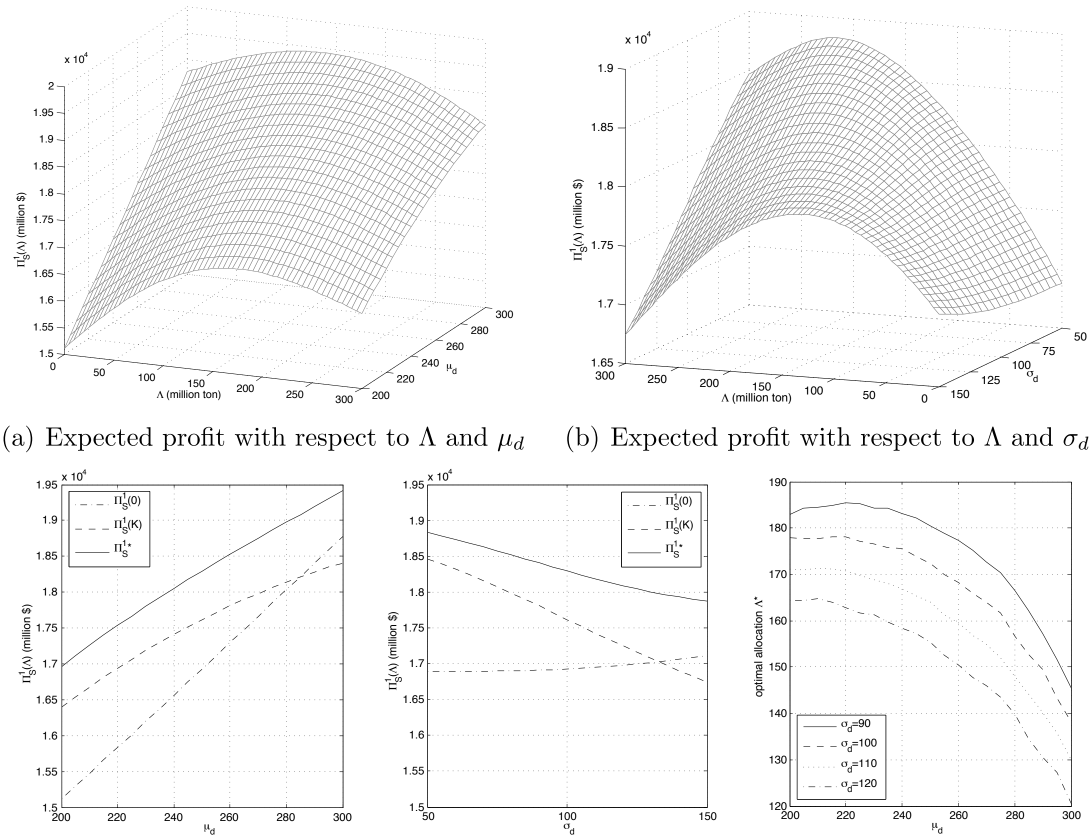
(c) Profit Comparison with Respect to $\alpha$

(d) Profit Comparison with Respect to $\beta$

(e) Optimal Allocation with Respect to $\alpha$ and $\beta$

Figure 4.11: Impact of $\alpha$ ($\gamma$) and $\beta$ ($\delta$) on Expected Profit and Optimal Allocation Decision

$\alpha$ and decreasing in $\beta$. The profit gap between the optimal allocation strategy and the total-contract strategy is increasing in $\alpha$ and decreasing in $\beta$, with average profit improvements of 3.84% and 3.88%, respectively. The profit gap between the optimal allocation strategy and the total-spot strategy is decreasing in $\alpha$ and increasing in $\beta$, with average profit improvements of 8.09% and 23.06%, respectively. Panel (e) further shows that the optimal contract allocation quantity $\Lambda^*$ is decreasing in $\alpha$ and increasing in $\beta$. This is to say the supplier should rely more on the spot channel if the innate supply level ($-\alpha$) in the spot market is low or the innate demand level ($\gamma$) is high, or if the demand and supply curves have low price elasticities ($\beta$, $\delta$).




(a) Expected profit with respect to $\Lambda$ and $\mu_d$

(b) Expected profit with respect to $\Lambda$ and $\sigma_d$

(c) Profit Comparison with Respect to $\mu_d$

(d) Profit Comparison with Respect to $\sigma_d$

(e) Optimal Allocation with Respect to $\mu_d$ and $\sigma_d$

Figure 4.12: Impact of $\mu_d$ and $\sigma_d$ on Expected Profit and Optimal Allocation Decision

### The Impact of $\mu_d$ and $\sigma_d$

Finally, Figure 4.12 shows the impact of the mean demand $\mu_d$ and the standard deviation $\sigma_d$ on the supplier's expected profit and his optimal first stage allocation decision. Panels (a) and (b) provide a 3-D overview. Panels (c) and (d) demonstrate that: The optimal expected profit is increasing in $\mu_d$ and decreasing in $\sigma_d$. The profit gap between the optimal allocation strategy and the total-contract strategy is increasing in both $\mu_d$ and $\sigma_d$, with average profit improvements of 4.07% and 4.03%, respectively. The profit gap between the optimal allocation strategy and the total-spot strategy is decreasing in both $\mu_d$ and $\sigma_d$, with average profit improvements of 7.98% and 8.05%, respectively. Panel (e) further shows that the optimal contract allocation quantity $\Lambda^*$ is decreasing in $\sigma$, and potentially unimodal in $\mu_d$. However, when the demand variability is high, the supplier should allocate more capacity to the spot market in the first stage as $\mu_d$ increases.

## 4.6. Conclusion

In this chapter, we looked at a commodity trading problem in which a commodity supplier such as Rio Tinto tries to decide the optimal allocation of his production capacity between a fixed-price contract channel and a spot market channel in order to maximize his total sales profit from the two. We discussed two different model settings in which the spot market is either *open*, i.e., the spot price is exogenous given by a random distribution, or *closed*, i.e., the spot price is endogenously determined by the supplier's quantity decision. We identified the supplier's optimal allocation policy under both circumstances and demonstrated how the optimal decision changes with respect to the key model parameters. We further quantified the average profit improvement of adopting a mixed-channel strategy versus using a single contract channel or a single spot channel through numerical analysis.

For the open spot market scenario, we found that the demand-price correlation and a risk-averse attitude are two reasons for the supplier to adopt a dual-channel strategy. In particular, the supplier should allocate more quantity to the spot channel if the contract channel demand and the spot price are more positively correlated, and he should allocate more to the contract channel if he is more risk-averse. Additionally, we ascertain that the optimal quantity allocated to the spot market should also increase in the average spot price, the demand variability, or the spot price variability. It should decrease in the average contract channel demand.

For the closed spot market case, we believe that we are the first to explicitly model the phenomenon in which the commodity supplier's single quantity decision affects both the demand curve and the supply curve in the spot market and further determines the equilibrium spot price. In addition to the detailed analysis in the main sections, we can extract the following managerial insights, which are instructive to a large commodity supplier such as Rio Tinto: First, the stronger the large supplier's market-leading impact or the weaker the resource competition between the large and the small suppliers (denoted by an increasing $\Omega$), the less profit the supplier can obtain from the spot market. Accordingly, the supplier should rely more on the contract channel in the first stage. Second, the more market options the buyers have during the second stage (denoted by a decreasing $\theta$), the less navigating power (in terms of shifting the demand curve) the supplier owns in the spot market; and thus the supplier should rely more on the contract channel in the first stage as well. Third, if the variability of the contract channel demand is high, then the supplier should reserve more capacity for the spot market in the first stage as the average contract demand increases, with the hope of driving more demand to the spot channel and creating a potentially high spot price.

In sum, we believe that multi-channel commodity trading is a very important practice in the modern global economy and many opportunities for economic and operational investigation of this topic still exist. We want to point out that although this research is motivated by the business practice in the iron ore industry, the modeling methodology and managerial insights are certainly applicable to other industries with similar channel choices to make.

# Chapter 5

# Conclusions

## 5.1. Major Results and Contributions

In this dissertation, we investigated how firms can utilize dual-channel sourcing and selling strategies to reduce their capacity procurement inefficiency and maximize their total sales income when faced with an increasingly uncertain business environment. We provide a detailed recap of the main discussions below.

In Chapter 2, we studied from a theoretical perspective a dynamic dual-source capacity expansion problem with backorders and demand forecast updates, in which a capital-intensive firm procures production capacity from both a flexible (fast) source and a base (slow) source. Assuming that the forecast updates follow an additive MMFE process and that the two capacity sources have consecutive zero-one lead-times, we formulated a dynamic programming recursion with two state variables: the capacity position (on-hand plus on-order) and the modified backorder level (actual backorder plus initial market information), and two decision variables: the expand-to capacity positions after ordering from the flexible source and the base source, respectively. We demonstrated joint concavity for the objective function and showed that the base orders follow a state-dependent base-stock policy. However, an optimal base-stock policy does not exist for the flexible source, and the flexible orders only follow a partial-base-stock policy. We further established some monotonicity properties for the (partial) base-stock levels, and quantified the value of having dual capacity sources and demand forecast updates using numerical analysis. We also investigated a brief extension of the model in which inventory and capacity decisions must be made jointly. We identified that once the capacity decisions are made, the production level can be determined according to a modified-base-stock policy, and that during each period it is optimal to order capacity from the flexible source only if all capacity after expansion would be used to produce inventory.

Our work in Chapter 2 complemented the extant OM literature on firms' multi-channel procurement strategies. The results showed that the dual-source capacity expansion problem with demand backlogging is categorically different from and more complicated than both its inventory counterpart and the dual-source capacity expansion problem with lost sales. Therefore, we called for heuristic solutions that firms can easily implement in practice to procure production or service capacity from multiple sources, the leadtimes of which are likely to be nonconsecutive.

Chapter 3 is thus a direct response to this call. In this chapter, we constructed a dual-mode equipment procurement heuristic (DMEP) to help Intel, the leading semiconductor manufacturer, improve its capital equipment procurement practice. DMEP enables the firm to dynamically order production capacity from two complementary service modes of an equipment supplier following a forecast revision process, during which the firm constantly adjusts its forecast for both the mean and the variance associated with future periods' demand. The entire DMEP heuristic consists of three layers. At the execution layer, we developed a rolling-horizon algorithm which allows the firm to solve a stochastic program during each period to determine the current-period order quantities through both supply modes, taking into consideration the revised demand information and subject to a reservation quantity constraint. At the reservation layer, before the planning horizon starts, the heuristic enumerates a large number of possible evolution paths of the initial forecasting profile, and determines the optimal reservation levels with both supply modes using a sample average approximation method. At the contract negotiation layer, the heuristic adopts an efficient-frontier approach and produces iso-profit curves in a leadtime-price quadrant based on sensitivity analyses. When the firm is about to negotiate the (*leaditme, price*) terms for both service modes with the equipment supplier, he can then conveniently compare different contract terms with the help of these iso-profit curves. We then demonstrated through actual numerical examples how DMEP can be used as an effective decision-support tool at Intel to manage the capital procurement process. We observed that the flexible mode is used under either a mean forecast shock or a demand realization shock; the firm tends to rely more on the flexible mode when the demand forecast uncertainty or the service level target is higher, or when the firm is more risk-averse.

The DMEP heuristic we constructed in this chapter is a fairly efficient tool that firms in capital-intensive industries can adopt to reduce their equipment procurement costs while still maintaining high service levels. To the best of our knowledge, we are the first to provide such a holistic solution to cover the operational, tactical, and strategic capacity decision problems. Furthermore, with slight changes of the state transition equations, DMEP can be easily adapted to address multi-source inventory control problems as well.

In Chapter 4, we switched our attention from reducing costs to boosting revenues. We looked at a commodity trading problem and investigated how a commodity supplier can strategically allocate his production capacity between a fixed-price contract channel and an uncertain spot channel, with the purpose of maximizing the total sales revenue from the two. We first looked at a scenario in which the spot market is open and the spot price is given by a random distribution. We showed that if the contract channel demand is certainly higher than the supplier's capacity, then the supplier may adopt a dual-channel allocation policy only if he is risk-averse and the average spot price is moderately higher than the fixed contract price. If the contract channel demand can be lower than the supplier's capacity and follows a bivariate normal distribution together with the spot price, then the supplier is better off adopting a dual-channel allocation strategy even when he is risk-neutral, as long as the contract demand and the spot price are positively correlated. We then looked at a case where the spot market is closed and the spot price is jointly determined by the demand and supply curves in the market, both of which can be affected by the supplier's allocation decision. We demonstrated that the supplier should adopt a single-channel strategy (either total-contract or total-spot) if the spot demand curve is endogenous while the spot supply curve is exogenous. If both curves are endogenously affected by the supplier's allocation decision, however, the supplier's expected profit function is convex-concave and a dual-channel strategy can be optimal, as long as the shifting effect of the supplier's spot allocation quantity on the default spot supply curve is stronger than shifting effect of the switched-over contract demand on the default spot demand curve. We also discussed how the optimal allocation quantity would change with respect to key model parameters and studied two special cases where the contract channel demand is either deterministic or two-point distributed. We further carried out a comprehensive numerical analysis to quantify the value of using dual-channels and the sensitivity of the optimal decisions.

Global commodity trading occurs at higher frequencies and larger scales as the economic activities of different countries become more interconnected. The capacity allocation model we constructed in this chapter is therefore instructive to help commodity suppliers take advantage of multiple sales channels to maximize their total income through the trading process. Decision-makers can use our model to determine the effect of key production and market parameters on the optimal allocation quantities to the fixed-price channel and the spot channel, respectively. The modeling methodology and managerial insights are also applicable to other industries with similar channel choices to make.

## 5.2. Directions for Future Extensions

For the dual-mode equipment procurement problem studied in Chapter 3, we ignored the impact of inventory carry-over on firms' capacity decisions. However, we would like to note that the execution-level algorithm can be modified to include the option of holding inventory easily. One must define a decision variable for inventory for each period and parameters for inventory holding cost and salvage value. As a result, the execution-level problem would be slightly more complicated since, in addition to the base and flexible capacity execution levels, the optimization would also need to calculate the optimal inventory levels as well. Our preliminary numerical studies show that, when holding inventory is an option, the firm tends to carry inventory and (1) build up capacity earlier, (2) order less capacity, and (3) decrease the percentage of flexible capacity reserved and exercised. It would be worthwhile to provide some theoretical justifications to these findings.

For the strategic capacity allocation problem investigated in Chapter 4, one could internalize the forward contract price negotiation between the supplier and the buyers using either a Stackelberg game or a Nash bargaining process. Alternatively, one could investigate a multi-period problem in which the current period's contract price is determined based on last period's realized spot price. It would also be meaningful to incorporate a detailed analysis for the buyer's side, and discuss how the commodity supplier's capacity allocation strategy would affect the total supply chain performance.

Furthermore, it could be interesting to combine the multi-channel sourcing problem with the multi-channel distribution problem, and study a setting in which the firm has a wide range of leverage from both the supply side and the demand side to adapt to the changing market conditions. We leave these opportunities of further analysis to other researchers.

# Appendix A

# Supplementary Discussion for Chapters 2 and 3

## A.1. The Case of Inventory Carry-Over

One assumption in Chapter 2's discussion is that the firm's production quantity during each period cannot be carried-over to the next period, and thus the inventory decision problem is eliminated from our consideration. This is indeed the situation for many service agencies with customized products, such as call centers, or for firms producing perishable goods. Some firms in the electronic industry also adopt a build-to-order policy and strive to keep its inventory level as low as possible in order to minimize the procurement cost. However, because of the production leadtime, many other firms would prefer to build to stock, under which circumstance capacity and inventory decisions need to be made jointly. In a related work, Angelus and Porteus (2002) address the problem of simultaneous production and capacity management under stochastic demand for produce-to-stock goods. They investigate a single-sourcing case and establish a target interval policy for capacity planning: it is optimal to make the smallest necessary change to bring the production capacity into a given target interval. Below, we extend their model to a dual-sourcing case by adding an inventory layer to our previous dual-source capacity expansion problem.

We investigate a case where both capacity and inventory are built to stock, i.e., need to be determined before all the demand randomness realizes. In the following

DP recursion, the first state variable $x$ still denotes the firm's modified capacity level (on-hand plus on-order); the second state variable $y$ here represents the firm's inventory position, for which a positive value means inventory on hand and a negative value implies backorders. To simplify notation, we eliminate $\mu_n$ and assume the final market information $\varepsilon_n^2$ has a mean of $\mu_n$; and that $\varepsilon_n^2 > 0$ holds almost surely.

$$\begin{aligned} J_n(x,y) \;=\; & \max_{x\le x'\le x'';y\le z\le y+x'} \mathbb{E}\,\{p_n[\min(\varepsilon_n^2, z^+) + z^- - y^-] - c_b(x''-x') - c_f(x'-x) \\ & - c_h x' - c_z(z-y) + \delta J_{n+1}(x'', z-\varepsilon_n^2 - \varepsilon_{n+1}^1)\} \qquad \text{(A.1.1)} \end{aligned}$$

for $n = 0, 1, \cdots, N$; $\varepsilon_{N+1}^2 = 0$ and $J_{N+1}(x,y) = c_u y$. Also notice that in the above formula, $z^+ = \max(z,0)$ and $z^- = \min(z,0)$. More explicitly, Equation (A.1.1) can be rewritten as

$$J_n(x,y) = \max_{x\le x'\le x''} \Big[v_n(x,x',x'',y) - \gamma(x,x',x'',y)\Big], \qquad \text{(A.1.2)}$$

where $\gamma(x,x',x'',y) = c_b(x''-x') + c_f(x'-x) + c_h x' - c_z y$, and

$$v_n(x,x',x'',y) = \max_{y\le z\le y+x'} \begin{cases} \mathbb{E}\,\{p_n \min(\varepsilon_n^2, z) - c_z z + J_{n+1}(x'', z-\varepsilon_n^2-\varepsilon_{n+1}^1)\}, & \text{for } y \ge 0 \\ \mathbb{E}\,\{p_n(z-y) - c_z z + J_{n+1}(x'', z-\varepsilon_n^2-\varepsilon_{n+1}^1)\}, & \text{for } y<0,\ x'<|y| \\ g_n(x'',z,y), & \text{for } y<0,\ x'\ge|y| \end{cases} \qquad \text{(A.1.3)}$$

$$g_n(x'',z,y) = \begin{cases} \mathbb{E}\,\{p_n(\min(\varepsilon_n^2, z) - y) - c_z z + J_{n+1}(x'', z-\varepsilon_n^2-\varepsilon_{n+1}^1)\}, & \text{for } 0\le z\le y+x' \\ \mathbb{E}\,\{p_n(z-y) - c_z z + J_{n+1}(x'', z-\varepsilon_n^2-\varepsilon_{n+1}^1)\}, & \text{for } y\le z<0 \end{cases} \qquad \text{(A.1.4)}$$

In the following part, we investigate some analytical properties of the optimal inventory and capacity policies. Note that the scope of our discussion is rather limited due to the complexity of the model.

**Proposition A.1.1.** *Both $J_n(x, y)$ and $v_n(x, x', x'', y)$ are concave; the objective function of $v_n(x, x', x'', y)$ is concave in the production decision $z$ with an unconstrained maximizer $\hat{z}(x'')$; and the optimal inventory decision $z^*$ is captured by a modified base-stock policy: build as close to $\hat{z}(x'')$ as possible in the region $[y, y + x']$.*

Proposition A.1.1 shows that once the capacity decisions are made, the inventory decision would follow a well-behaved base-stock type of policy. To explicitly describe the optimal base-stock level $\hat{z}$, however, one needs to rely on numerical methods.

**Proposition A.1.2.** *Given $c_f > c_b$, at optimality we have $x' > x$ only if $z^* = y + x'$.*

The above proposition says that when both capacity and inventory are built to stock, the firm orders capacity from the flexible mode during a certain period only if all capacity after expansion will be used to produce inventory. The converse is not necessarily true though.

## A.2. A High-Level Discussion on Risk Aversion

Here we explore some general analytical results associated with risk-averse decision-making. The purpose of the subsequent general discussion is two-fold: 1. It generates additional theoretical contribution to the OM literature. 2. The entire DMEP heuristic we construct is very complex and intertwined, and it would be a great computational challenge to directly address all the modeling specifics in a theoretical exploration; hence we hope that a high-level analytical discussion can at least provide some justification for our numerical observations in Section 3.4.3 with regard to risk-averse decision-making.

The following proposition explains how a concave increasing utility function affects an individual or a company's optimal utility-maximizing decision in a stochastic context, given different properties that the original objective function possesses.

**Proposition A.2.1.** *Let $\zeta$ be a random variable. Assume that a continuous differentiable function $f(x, \zeta)$ is concave in $x$, and that a continuous differentiable function $g(\cdot)$ is concave and increasing. Let $x^* = \arg\max_x \mathbb{E}_\zeta f(x, \zeta)$ and $\hat{x}^* = \arg\max_x \mathbb{E}_\zeta g(f(x, \zeta))$, we have:*

Table A.1: Change of Optimal Solution under Concave Increasing Utility Function

| Case | Sufficient conditions on $f(x,\zeta)$ in a neighborhood of $x^*$ | | Conclusion |
|---|---|---|---|
| $(i)$ | submodular in $(x,\zeta)$; | increasing in $\zeta$ | $\hat{x}^* \geq x^*$ |
| $(ii)$ | supermodular in $(x,\zeta)$; | decreasing in $\zeta$ | |
| $(iii)$ | submodular in $(x,\zeta)$; | decreasing in $\zeta$ | $\hat{x}^* \leq x^*$ |
| $(iv)$ | supermodular in $(x,\zeta)$; | increasing in $\zeta$ | |

The additional properties we imposed on the original objective function $f(x,\zeta)$ are only sufficient conditions. One may argue that the combination of modularity[1] and monotonicity is too strong a condition, but it turns out that these conditions only need to be satisfied in a small neighborhood of $x^*$ – the original maximizer of $\mathbb{E}f(x,\zeta)$ before $g(\cdot)$ is applied. For the more general case where $f(x,\zeta)$ observes modularity but not monotonicity, as illustrated by Figure A.1(a), similar conclusions can be made if $g(\cdot)$ satisfies certain properties as described by Corollary A.2.2 below.

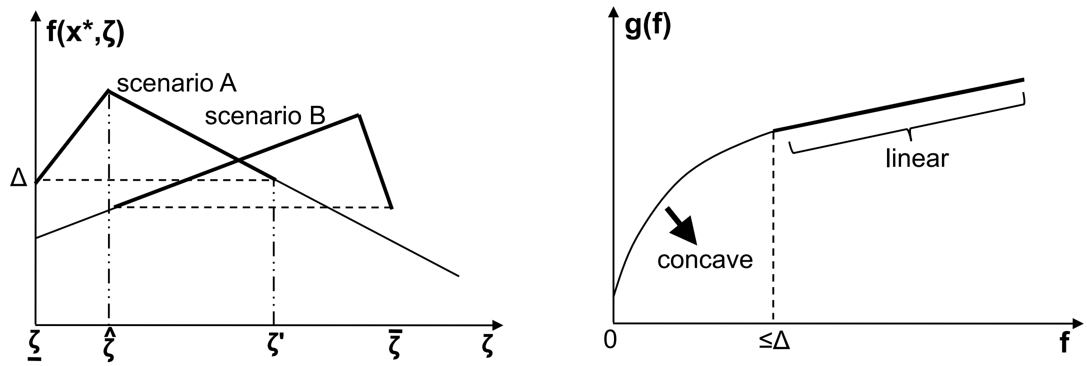
**Corollary A.2.2.** *Assume $\zeta$ has a bounded support on $[\underline{\zeta},\bar{\zeta}]$, $f(x,\zeta)$ is continuous differentiable and concave in $x$, $f(x^*,\zeta)$ is unimodal in $\zeta$. Also assume that the continuous differentiable function $g(f)$ is concave increasing for $f < \Delta$ and linear increasing for $f \geq \Delta$, where $\Delta = \max(f(x^*,\underline{\zeta}), f(x^*,\bar{\zeta}))$. Then we have:*

Table A.2: Change of Optimal Solution under Concave Increasing Utility Function

| Case | Sufficient conditions on $f(x,\zeta)$ in a neighborhood of $x^*$ | | Conclusion |
|---|---|---|---|
| $(v)$ | submodular in $(x,\zeta)$; | $f(x^*,\underline{\zeta}) < f(x^*,\bar{\zeta})$ | $\hat{x}^* \geq x^*$ |
| $(vi)$ | supermodular in $(x,\zeta)$; | $f(x^*,\underline{\zeta}) \geq f(x^*,\bar{\zeta})$ | |
| $(vii)$ | submodular in $(x,\zeta)$; | $f(x^*,\underline{\zeta}) \geq f(x^*,\bar{\zeta})$ | $\hat{x}^* \leq x^*$ |
| $(viii)$ | supermodular in $(x,\zeta)$; | $f(x^*,\underline{\zeta}) < f(x^*,\bar{\zeta})$ | |

In a real business context such as a profit-maximizing newsvendor setting, the above segmented utility function $g(\cdot)$ is actually a reasonable one. It simply says that

[1]We use "modularity" as a general reference for both submodularity and supermodularity.




(a) Two scenarios of a unimodal $f(x^*, \zeta)$ in $\zeta$

(b) A particular type of function $g$

Figure A.1: The Case in Which $f(x^*, \zeta)$ Is Unimodal in $\zeta$

the firm is risk neutral when the monetary income is high (higher than $\Delta$ in this case), while it tends to be risk-averse when the monetary income is low. In other words, the firm is less risk-averse when it becomes richer, which is to some extent consistent with the property of decreasing absolute risk aversion.

Directly applying the above general analysis to our comprehensive model in Section 3.4.3, however, is quite difficult. First, in our model both the decision $x$ and the random factor $\zeta$ are multi-dimension vectors: $x$ refers to $(B^T, F^T)$, and $\zeta$ refers to the entire mean forecast evolution space $M$. Second, $\mathcal{F}$ itself is the value function of a constrained stochastic optimization, and potentially all the derivative investigation involves langrangian formulation. Therefore, in the dissertation we resort to numerical analysis to evaluate the impact of risk aversion.

# Appendix B

# Proofs

**Proof of Lemma 2.4.1**: We prove the lemma using an inductive argument. At the final stage $N$, we have $\tilde{J}_N(\tilde{x}_N, \tilde{y}_N) + G_N(\tilde{y}_N)$

$$= \max_{\tilde{x}_N \le \tilde{x}'_N \le \tilde{x}''_N} \mathbb{E}\Big[ -(p_N - 0)(\tilde{y}_N + \varepsilon_N^2 + \mu_N - \tilde{x}'_N)^+ - c_f(\tilde{x}'_N - \tilde{x}_N) - c_b(\tilde{x}''_N - \tilde{x}'_N) - c_h\tilde{x}'_N - \delta c_u(\tilde{y}_N + \varepsilon_N^2 + \mu_N - \tilde{x}'_N)^+\Big] + p_N\tilde{y}_N + \mathbb{E}p_N(\varepsilon_N^2 + \mu_N)$$

$$= \max_{\tilde{x}_N \le \tilde{x}'_N \le \tilde{x}''_N} \mathbb{E}\Big[p_N \min\{\tilde{y}_N + \varepsilon_N^2 + \mu_N, \tilde{x}'_N\} - c_f(\tilde{x}'_N - \tilde{x}_N) - c_b(\tilde{x}''_N - \tilde{x}'_N) - c_h\tilde{x}'_N - \delta c_u(\tilde{y}_N + \varepsilon_N^2 + \mu_N - \tilde{x}'_N)^+\Big]$$

$$= J_N(\tilde{x}_N, \tilde{y}_N).$$

Now, assume the relation holds for period $n+1$, $n < N$; that is,

$$\tilde{J}_{n+1}(\tilde{x}_{n+1}, \tilde{y}_{n+1}) + G_{n+1}(\tilde{y}_{n+1}) = J_{n+1}(\tilde{x}_{n+1}, \tilde{y}_{n+1}).$$

Then $\tilde{J}_n(\tilde{x}_n, \tilde{y}_n) + G_n(\tilde{y}_n)$

$$= \max_{\tilde{x}_n \le \tilde{x}'_n \le \tilde{x}''_n} \mathbb{E}\Big[ -(p_n - \delta p_{n+1})(\tilde{y}_n + \varepsilon_n^2 + \mu_n - \tilde{x}'_n)^+ - c_f(\tilde{x}'_n - \tilde{x}_n) - c_b(\tilde{x}''_n - \tilde{x}'_n) - c_h\tilde{x}'_n + \delta[J_{n+1}(\tilde{x}''_n, (\tilde{y}_n + \varepsilon_n^2 + \mu_n - \tilde{x}'_n)^+ + \varepsilon_{n+1}^1) - G_{n+1}((\tilde{y}_n + \varepsilon_n^2 + \mu_n - \tilde{x}'_n)^+ + \varepsilon_{n+1}^1)]\Big] + G_n(\tilde{y}_n)$$

$$
\begin{aligned}
= \max_{\tilde{x}_n \le \tilde{x}'_n \le \tilde{x}''_n} \quad & \mathbb{E}\Big[ -(p_n - \delta p_{n+1})(\tilde{y}_n + \varepsilon_n^2 + \mu_n - \tilde{x}'_n)^+ - c_f(\tilde{x}'_n - \tilde{x}_n) - c_b(\tilde{x}''_n - \tilde{x}'_n) - c_h \tilde{x}'_n \\
& + \delta J_{n+1}(\tilde{x}''_n, (\tilde{y}_n + \varepsilon_n^2 + \mu_n - \tilde{x}'_n)^+ + \varepsilon_{n+1}^1) - \delta p_{n+1}((\tilde{y}_n + \varepsilon_n^2 + \mu_n - \tilde{x}'_n)^+ \\
& + \varepsilon_{n+1}^1) - \delta \sum_{k=n+2}^{N} \delta^{k-(n+1)} p_k \varepsilon_k^1 - \delta \sum_{k=n+1}^{N} \delta^{k-(n+1)} p_k (\varepsilon_k^2 + \mu_k) + p_n \tilde{y}_n \\
& + \sum_{k=n+1}^{N} \delta^{k-n} p_k \varepsilon_k^1 + \sum_{k=n}^{N} \delta^{k-n} p_k (\varepsilon_k^2 + \mu_k) \Big] \\
= \max_{\tilde{x}_n \le \tilde{x}'_n \le \tilde{x}''_n} \quad & \mathbb{E}\Big[ -p_n(\tilde{y}_n + \varepsilon_n^2 + \mu_n - \tilde{x}'_n)^+ + p_n(\tilde{y}_n + \varepsilon_n^2 + \mu_n) - c_f(\tilde{x}'_n - \tilde{x}_n) - c_b(\tilde{x}''_n - \tilde{x}'_n) \\
& - c_h \tilde{x}'_n + \delta J_{n+1}(\tilde{x}''_n, (\tilde{y}_n + \varepsilon_n^2 + \mu_n - \tilde{x}'_n)^+ + \varepsilon_{n+1}^1) \Big] \\
= J_n(\tilde{x}_n, \tilde{y}_n). &
\end{aligned}
$$

□

**Proof of Lemma 2.4.2:** Trivially, $\tilde{J}_N(\tilde{x}_N, \tilde{y}_N)$ is decreasing in $\tilde{y}_N$. Let $\alpha \in [0,1]$ and $\bar{\alpha} = 1 - \alpha$. Given $(\tilde{x}'_{N,1}, \tilde{y}_{N,1})$ and $(\tilde{x}'_{N,2}, \tilde{y}_{N,2})$, since

$$
\begin{aligned}
& \alpha(\tilde{y}_{N,1} + \varepsilon_N^2 + \mu_N - \tilde{x}'_{N,1})^+ + \bar{\alpha}(\tilde{y}_{N,2} + \varepsilon_N^2 + \mu_N - \tilde{x}'_{N,2})^+ \\
& \quad \ge ((\alpha \tilde{y}_{N,1} + \bar{\alpha} \tilde{y}_{N,2}) + \varepsilon_N^2 + \mu_N - (\alpha \tilde{x}'_{N,1} + \bar{\alpha} \tilde{x}'_{N,2}))^+,
\end{aligned}
$$

the objective function is concave in $(\tilde{x}_N, \tilde{y}_N, \tilde{x}'_N, \tilde{x}''_N)$. By concavity preservation under maximization, $\tilde{J}_N(\tilde{x}_N, \tilde{y}_N)$ is concave in $(\tilde{x}_N, \tilde{y}_N)$.

Assuming $\tilde{J}_{n+1}(\tilde{x}_{n+1}, \tilde{y}_{n+1})$ is decreasing in $\tilde{y}_{n+1}$ and concave in $(\tilde{x}_{n+1}, \tilde{y}_{n+1})$, $n < N$, and given $(\tilde{x}_{n,1}, \tilde{y}_{n,1}, \tilde{x}'_{n,1}, \tilde{x}''_{n,1})$ and $(\tilde{x}_{n,2}, \tilde{y}_{n,2}, \tilde{x}'_{n,2}, \tilde{x}''_{n,2})$, we then have

$$
\begin{aligned}
& \alpha \tilde{J}_{n+1}(\tilde{x}''_{n,1}, (\tilde{y}_{n,1} + \varepsilon_n^2 + \mu_n - \tilde{x}'_{n,1})^+ + \varepsilon_{n+1}^1) + \bar{\alpha} \tilde{J}_{n+1}(\tilde{x}''_{n,2}, (\tilde{y}_{n,2} + \varepsilon_n^2 + \mu_n - \tilde{x}'_{n,2})^+ + \varepsilon_{n+1}^1) \\
\le \; & \tilde{J}_{n+1}(\alpha \tilde{x}''_{n,1} + \bar{\alpha} \tilde{x}''_{n,2}, \alpha(\tilde{y}_{n,1} + \varepsilon_n^2 + \mu_n - \tilde{x}'_{n,1})^+ + \bar{\alpha}(\tilde{y}_{n,2} + \varepsilon_n^2 + \mu_n - \tilde{x}'_{n,2})^+ + \varepsilon_{n+1}^1) \\
\le \; & \tilde{J}_{n+1}(\alpha \tilde{x}''_{n,1} + \bar{\alpha} \tilde{x}''_{n,2}, ((\alpha \tilde{y}_{n,1} + \bar{\alpha} \tilde{y}_{n,2}) + \varepsilon_n^2 + \mu_n - (\alpha \tilde{x}'_{n,1} + \bar{\alpha} \tilde{x}'_{n,2}))^+ + \varepsilon_{n+1}^1).
\end{aligned}
$$

The first inequality follows from the joint concavity of $\tilde{J}_{n+1}(\tilde{x}_{n+1}, \tilde{y}_{n+1})$; the second inequality is due to $\tilde{J}_{n+1}(\tilde{x}_{n+1}, \tilde{y}_{n+1})$ being decreasing in its second argument and the fact that $\alpha u^+ + (1-\alpha)v^+ \ge (\alpha u + (1-\alpha)v)^+$. Thus, we can claim that

$J_{n+1}(\tilde{x}_n'', (\tilde{y}_n + \varepsilon_n^2 + \mu_n - \tilde{x}_n')^+ + \varepsilon_{n+1}^1)$ is concave in $(\tilde{y}_n, \tilde{x}_n', \tilde{x}_n'')$, hence trivially concave in $(\tilde{x}_n, \tilde{y}_n, \tilde{x}_n', \tilde{x}_n'')$ (since it does not contain $\tilde{x}_n$). In the objective function of recursion (2.4.2), all terms are concave in $(\tilde{x}_n, \tilde{y}_n, \tilde{x}_n', \tilde{x}_n'')$, so the expectation is as well. Applying the concavity preservation theorem under maximization (Topkis, 1978:314), we conclude that the value function $\tilde{J}_n(\tilde{x}_n, \tilde{y}_n)$ is concave in $(\tilde{x}_n, \tilde{y}_n)$. Also, due to the fact that $p_n > \delta p_{n+1}$, we have that $\tilde{J}_n(\tilde{x}_n, \tilde{y}_n)$ decreases in $\tilde{y}_n$. This completes the induction. □

**Proof of Proposition 2.4.3:** The result follows directly from Lemma 2.4.1 and Lemma 2.4.2. □

**Proof of Propositions 2.4.4-2.4.6:** By Proposition 2.4.3, we know that $V_n(\tilde{x}_n, \tilde{y}_n, \tilde{x}_n', \tilde{x}_n'')$ is concave in $(\tilde{x}_n', \tilde{x}_n'')$. Thus, we can define $S_n^F$ and $S_n^B$ (we temporarily suppress the state parameter $\tilde{y}_n$ for expositional simplicity) as:

$$(S_n^F, S_n^B) = \arg \max_{0 \le \tilde{x}_n' \le \tilde{x}_n''} V_n(\tilde{x}_n, \tilde{y}_n, \tilde{x}_n', \tilde{x}_n''). \tag{B.0.1}$$

The claim that $S_n^F \le S_n^B$ follows directly from the constraint that $\tilde{x}_n' \le \tilde{x}_n''$. When $\tilde{x}_n \le S_n^F$, the optimal expand-to capacity levels $(\tilde{x}_n', \tilde{x}_n'')$ are equal to $(S_n^F, S_n^B)$. For $\tilde{x}_n > S_n^F$, we show $\tilde{x}_n' = \tilde{x}_n$ via a contradiction argument. Assume that $(\kappa_n', \tilde{x}_n'')$ are the optimal expand-to capacity positions where $\tilde{x}_n'' \ge \kappa_n' > \tilde{x}_n$. We must have

$$V_n(\tilde{x}_n, \tilde{y}_n, \kappa_n', \tilde{x}_n'') \le V_n(\tilde{x}_n, \tilde{y}_n, S_n^F, S_n^B)$$

due to the global optimality of $(S_n^F, S_n^B)$. Also, since $S_n^F < \tilde{x}_n < \kappa_n'$, there exists some $\theta \in [0, 1]$ with $\bar{\theta} = 1 - \theta$, such that $\tilde{x}_n = \theta S_n^F + \bar{\theta}\kappa_n'$. Further letting $\kappa_n'' = \theta S_n^B + \bar{\theta}\tilde{x}_n''$, we have

$$\begin{aligned} V_n(\tilde{x}_n, \tilde{y}_n, \tilde{x}_n, \kappa_n'') &= V_n(\tilde{x}_n, \tilde{y}_n, \theta S_n^F + \bar{\theta}\kappa_n', \theta S_n^B + \bar{\theta}\tilde{x}_n'') \\ &\ge \theta V_n(\tilde{x}_n, \tilde{y}_n, S_n^F, S_n^B) + \bar{\theta} V_n(\tilde{x}_n, \tilde{y}_n, \kappa_n', \tilde{x}_n'') \\ &\ge V_n(\tilde{x}_n, \tilde{y}_n, \kappa_n', \tilde{x}_n''), \end{aligned}$$

which contradicts the fact that $(\kappa_n', \tilde{x}_n'')$ are the optimal expand-to capacity positions. Therefore, it must be the case that $\kappa_n' = \tilde{x}_n$, i.e., $\tilde{x}_n' = \tilde{x}_n$ for $\tilde{x}_n > S_n^F$. We conclude that a state-dependent base-stock policy is optimal for the flexible source.

For the base source, we have already shown that it is optimal to expand the capacity position to $S_n^B$ when $\tilde{x}_n \leq S_n^F$. Since demand is finite, there must exist a capacity value $\underline{S}_n^B \geq S_n^F$ such that no base orders will be placed when $\tilde{x}_n > \underline{S}_n^B$. For $\tilde{x}_n \in (S_n^F, \underline{S}_n^B]$, the following counterexample shows that the optimal expand-to capacity position may depend on both $\tilde{x}_n$ and $\tilde{y}_n$.

*Example.* Assume there are only two periods and that demand in each period is deterministic with value 30. $c_f$ is sufficiently large so that the base supplier is the only choice, with a leadtime of one period. If at the beginning of period 1 we have zero on-hand capacity, then, anticipating that all the demand in period 1 will be backlogged into the next period, we will expand the capacity position to 60 to satisfy the total demand of 60 units in period 2. Suppose instead at the beginning of period 1, there are 20 units of on-hand capacity. Then only 10 units of demand of period 1 will not be satisfied and hence will be backlogged into period 2, rendering period 2's total demand to 40 units. Given this, it is now optimal to order 20 units of capacity and expand the capacity position to 40, instead of 60 as in the previous scenario. We have demonstrated that in this case, the optimal expand-to capacity position is decreasing in the initial capacity position within a certain range. Hence, we prove that a base-stock policy cannot be optimal for the base source.

Now, it remains to show how the (partial) base-stock levels $S_n^F(\tilde{y}_n)$ and $S_n^B(\tilde{y}_n)$ are dependent on the state $\tilde{y}_n$. We rearrange Equation (2.4.2) as follows:

$$\max_{\tilde{x}_n \leq \tilde{x}'_n \leq \tilde{x}''_n} \mathbb{E} \Big[ -(p_n - \delta p_{n+1})(\tilde{y}_n - \tilde{x}'_n + \varepsilon_n^2 + \mu_n)^+ + (c_f - c_b + c_h)(\tilde{y}_n - \tilde{x}'_n) + c_f(\tilde{x}_n - \tilde{y}_n) - c_b(\tilde{x}''_n - \tilde{y}_n) - c_h \tilde{y}_n + \delta \tilde{J}_{n+1}(\tilde{x}''_n, (\tilde{y}_n - \tilde{x}'_n + \varepsilon_n^2 + \mu_n)^+ + \varepsilon_{n+1}^1) \Big]. \tag{B.0.2}$$

Also recall that

$$(S_n^F(0), S_n^B(0)) = \arg \max_{0 \leq \tilde{x}'_n \leq \tilde{x}''_n} V_n(\tilde{x}_n, 0, \tilde{x}'_n, \tilde{x}''_n).$$

Note that in Equation (B.0.2), $\tilde{y}_n - \tilde{x}'_n$ can be treated as one quantity. As we change $\tilde{y}_n$ by a certain amount, the optimal value of $\tilde{x}'_n$ in Equation (B.0.1), namely $S_n^F(\tilde{y}_n)$, should shift by exactly the same amount, as long as the newly reached value would not exceed the previous optimal value of $\tilde{x}''_n$. Within this range, $S^B_n(\tilde{y}_n)$ is independent of $\tilde{y}_n$. Hence, when $\tilde{y}_n \leq S^B_n(0) - S^F_n(0)$, we have $S^F_n(\tilde{y}_n) = S^F_n(0) + \tilde{y}_n$ and $S^B_n(\tilde{y}_n) = S^B_n(0)$. Via an analogous rearrangement for period $N$, $S^F_N(\tilde{y}_N) = S^F_N(0) + \tilde{y}_N$ always holds. □

**Proof of Proposition 2.4.7:** For the ease of analysis, we rewrite the DP recursion of the two-period problem in the following cascade form:

$$\tilde{J}_1(\tilde{x}_1, \tilde{y}_1) = \max_{\tilde{x}'_1 \geq \tilde{x}_1} g_1(\tilde{x}'_1, \tilde{x}_1, \tilde{y}_1), \tag{B.0.3}$$

$$g_1(\tilde{x}'_1, \tilde{x}_1, \tilde{y}_1) = -(p_1 - \delta p_2)\mathbb{E}(\tilde{y}_1 + \varepsilon^2_1 + \mu_1 - \tilde{x}'_1)^+ - (c_f + c_h - c_b)\tilde{x}'_1 + c_f\tilde{x}_1 + \max_{\tilde{x}''_1 \geq \tilde{x}'_1} \Gamma_1(\tilde{x}'_1, \tilde{x}''_1, \tilde{y}_1), \tag{B.0.4}$$

$$\Gamma_1(\tilde{x}'_1, \tilde{x}''_1, \tilde{y}_1) = \delta\mathbb{E}\tilde{J}_2(\tilde{x}''_1, (\tilde{y}_1 + \varepsilon^2_1 + \mu_1 - \tilde{x}'_1)^+ + \varepsilon^1_2) - c_b\tilde{x}''_1, \tag{B.0.5}$$

and finally,

$$\tilde{J}_2(\tilde{x}_2, \tilde{y}_2) = \max_{\tilde{x}'_2 \geq \tilde{x}_2} g_2(\tilde{x}'_2, \tilde{x}_2, \tilde{y}_2), \tag{B.0.6}$$

$$g_2(\tilde{x}'_2, \tilde{x}_2, \tilde{y}_2) = -(p_2 + \delta c_u)\mathbb{E}(\tilde{y}_2 + \varepsilon^2_2 + \mu_2 - \tilde{x}'_2)^+ - (c_f + c_u)\tilde{x}'_2 + c_f\tilde{x}_2. \tag{B.0.7}$$

Claim 1: $g_2(\tilde{x}'_2, \tilde{x}_2, \tilde{y}_2)$ is decreasing in $\tilde{y}_2$, concave and supermodular in $(\tilde{x}'_2, \tilde{y}_2)$. The decreasing property is obvious. To see concavity, note that the plus function $(\cdot)^+$ is convex, and the other terms are linear. To see supermodularity, we apply Topkis's Theorem by directly checking the cross-partials[1]:

$$g^{(1)}_2(\tilde{x}'_2, \tilde{x}_2, \tilde{y}_2) = (p_2 + \delta c_u)[1 - \Phi_{\varepsilon^2_2}(\tilde{x}'_2 - \mu_2 - \tilde{y}_2)] - (c_f + c_u) \tag{B.0.8}$$

Since the above partial derivative is increasing in $\tilde{y}_2$, supermodularity is verified.
Claim 2: $\tilde{J}_2(\tilde{x}_2, \tilde{y}_2)$ is decreasing in $\tilde{y}_2$, concave and supermodular in $(\tilde{x}_2, \tilde{y}_2)$. Again, it's trivial to show the decreasing property. Concavity also follows directly from the concavity preservation theorem under maximization. To see the supermodularity of $\tilde{J}_2$, let $S(\tilde{y}_2)$ represent the unconstrained maximizer of $g_2$ (since $\tilde{x}_2$ only appears in a linear term, it does not affect the optimal solution); we know $S(\tilde{y}_2)$ is

[1]Superscript $(k)$ denotes taking derivative with respect to the $k$th argument

increasing in $\tilde{y}_2$ due to the supermodularity of $g_2$. Therefore, we have

$$\tilde{J}_2(\tilde{x}_2, \tilde{y}_2) = \begin{cases} g_2(S(\tilde{y}_2), \tilde{x}_2, \tilde{y}_2), & \text{if } S(\tilde{y}_2) \geq \tilde{x}_2 \\ g_2(\tilde{x}_2, \tilde{x}_2, \tilde{y}_2), & \text{if } S(\tilde{y}_2) < \tilde{x}_2 \end{cases} \tag{B.0.9}$$

$$\tilde{J}_2^{(1)}(\tilde{x}_2, \tilde{y}_2) = \begin{cases} 0, & \text{if } S(\tilde{y}_2) \geq \tilde{x}_2 \\ g_2^{(1)}(\tilde{x}_2, \tilde{x}_2, \tilde{y}_2), & \text{if } S(\tilde{y}_2) < \tilde{x}_2 \end{cases} \tag{B.0.10}$$

We must show $\tilde{J}_2^{(1)}(\tilde{x}_2, \tilde{y}_2)$ is increasing in $\tilde{y}_2$. Because $g_2(\tilde{x}'_2, \tilde{x}_2, \tilde{y}_2)$ is supermodular in $(\tilde{x}'_2, \tilde{y}_2)$, the only unobvious case is when $\tilde{y}_2$ increases such that $S(\tilde{y}_2)$ crosses the line from below $\tilde{x}_2$ to above $\tilde{x}_2$. Hence, we must prove that $g_2^{(1)}(\tilde{x}_2, \tilde{x}_2, \tilde{y}_2) \leq 0$ for $S(\tilde{y}_2) < \tilde{x}_2$. This holds since $g_2$ is concave, $g_2^{(1)}(S(\tilde{y}_2), \tilde{x}_2, \tilde{y}_2) = 0$, and $\tilde{x}_2 > S(\tilde{y}_2)$.
Claim 3: $\Gamma_1(\tilde{x}'_1, \tilde{x}''_1, \tilde{y}_1)$ is concave, submodular in $(\tilde{x}'_1, \tilde{x}''_1)$, and supermodular in $(\tilde{x}''_1, \tilde{y}_1)$.
The concavity of $\Gamma_1$ can be shown following an argument similar to the proof of Lemma 2.4.2. To verify sub- and super-modularity, we apply both Topkis's Theorem and Fubini's Theorem:

$$\Gamma_1^{(2)}(\tilde{x}'_1, \tilde{x}''_1, \tilde{y}_1) = \delta \mathbb{E} \tilde{J}_2^{(1)}(\tilde{x}''_1, (\tilde{y}_1 + \varepsilon_1^2 + \mu_1 - \tilde{x}'_1)^+ + \varepsilon_2^1) - c_b, \tag{B.0.11}$$

which is decreasing in $\tilde{x}'_1$ and increasing in $\tilde{y}_1$ since $\tilde{J}_2$ is supermodular. Hence, $\Gamma_1(\tilde{x}'_1, \tilde{x}''_1, \tilde{y}_1)$ is submodular in $(\tilde{x}'_1, \tilde{x}''_1)$ and supermodular in $(\tilde{x}''_1, \tilde{y}_1)$. Therefore, the unconstrained optimal solution of $\Gamma_1$, i.e., the unconstrained optimal base expand-to capacity position $\tilde{x}''_1$ is decreasing in $\tilde{x}'_1$ and hence decreasing in $\tilde{x}_1$, since $\tilde{x}'_1 = \tilde{x}_1$ in this region. □

**Proof of Proposition 2.5.1:** Notice that any optimal solution to either one of the single-source capacity expansion problems must also be a feasible solution to the dual-source capacity expansion problem (setting all decisions associated with the unused supplier to be zero). □

**Proof of Proposition A.1.1:** We prove the concavity results using induction. $J_{N+1}(x, y) = c_u y$ is obviously concave since it is linear. Assume $J_{n+1}$ is concave for $n \leq N$, then $J_{n+1}(x'', z - \varepsilon_n^2 - \varepsilon_{n+1}^1)$ is concave in $(x'', z)$ hence concave in $z$. Since

$\min(\varepsilon_n^2, z)$ is also concave in $z$ and all the other linear terms do not affect concavity, the only unobvious case is the concavity of the piecewise function $g_n(x'', z, y)$ on the entire interval $z \in [y, y+x']$ for $y < 0$ and $x' \geq |y|$, i.e., whether $g_n(x'', z, y)$ is continuously differentiable at $z = 0$; but this is true since:

$$\begin{aligned} \lim_{z \to 0_+} \frac{\partial g_n(x'', z, y)}{\partial z} &= \lim_{z \to 0_+} p_n(1 - \Phi_{\varepsilon_n^2}(z)) - c_z + \frac{\partial \mathbb{E} J_{n+1}(x'', z - \varepsilon_n^2 - \varepsilon_{n+1}^1)}{\partial z}|_{z=0} \\ &= p_n - c_z + \frac{\partial \mathbb{E} J_{n+1}(x'', z - \varepsilon_n^2 - \varepsilon_{n+1}^1)}{\partial z}|_{z=0} \quad (\varepsilon_n^2 > 0 \ a.s.) \\ &= \lim_{z \to 0_-} \frac{\partial g_n(x'', z, y)}{\partial z}; \end{aligned} \tag{B.0.12}$$

$$\begin{aligned} \lim_{z \to 0_+} \frac{\partial g_n^2(x'', z, y)}{\partial z^2} &= \lim_{z \to 0_+} -p_n \phi_{\varepsilon_n^2}(z) + \frac{\partial \mathbb{E} J_{n+1}^2(x'', z - \varepsilon_n^2 - \varepsilon_{n+1}^1)}{\partial z^2}|_{z=0} \\ &= \frac{\partial \mathbb{E} J_{n+1}^2(x'', z - \varepsilon_n^2 - \varepsilon_{n+1}^1)}{\partial z^2}|_{z=0} \quad (\varepsilon_n^2 > 0 \ a.s.) \\ &= \lim_{z \to 0_-} \frac{\partial g_n^2(x'', z, y)}{\partial z^2}. \end{aligned} \tag{B.0.13}$$

Hence, the objective function of $v_n(x, x', x'', y)$ is concave; and because $y \leq z \leq y + x'$ is a convex set, we have $v_n$ is also concave due to the concavity preservation theorem under maximization. Similarly, since $v_n(x, x', x'', y)$ is concave, $\gamma(x, x', x'', y)$ is affine, and $x \leq x' \leq x''$ is a convex set, we conclude that $J_n(x, y)$ is concave, completing the induction. With concavity, the modified base-stock policy follows. □

**Proof of Proposition A.1.2:** Assume for the purpose of contradiction that when $x' > x$, we have $z < y + x'$. Then according to Equation (A.1.2), (A.1.3), and (A.1.4), through decreasing $x'$ by $\Delta$ while keeping $x''$ and $z$ unchanged, we can strictly increase the expected profit by $(c_f - c_b + c_h)\Delta > 0$. We would keep doing so until either $z = y + x'$, confirming the claim, or $x' = x$, violating the premise. □

**Proof of Proposition 3.3.1:** We claim that the equivalent linear program is given in the following format (subscript $j$ here represents the $j$-th monte carlo sample path; for ease of exhibition, we do not display the decision variables $\vec{s}_{1,\cdots,N;j}$ under the maximization operator):

$$
\begin{aligned}
\underset{\vec{B}_{1,\cdots,N};\ \vec{F}_{1,\cdots,N}}{\text{maximize}} \quad & \lim_{M\to\infty} \frac{1}{M} \sum_{j=1}^{M} \left[ \sum_{i=1}^{N} \delta^i \{p_i s_{i,j} - c_b B_i - c_f F_i - c_h k_i\} - \delta^{N+1} c_u d_{N+1,j}^{rem} \right] \\
\text{subject to} \quad & s_{i,j} \le k_i \quad \text{for } i = 1, \cdots, N;\ j = 1, \cdots, M \qquad \text{(B.0.14)} \\
& s_{i,j} \le d_{i,j} + d_{i,j}^{rem} \quad \text{for } i = 1, \cdots, N;\ j = 1, \cdots, M \qquad \text{(B.0.15)} \\
& d_{i,j}^{rem} = d_{i-1,j} + d_{i-1,j}^{rem} - s_{i-1,j}, \\
& \qquad \text{for } i = 1, \cdots, N;\ j = 1, \cdots, M;\ \text{with } d_{1,j}^{rem} = 0 \qquad \text{(B.0.16)} \\
& \text{and constraints (3.3.3), (3.3.4), (3.3.6)} - \text{(3.3.8)}
\end{aligned}
$$

Comparing the above linear program with the original stochastic program, we observe several differences: (*i*) We rewrite the objective function using the sample average approximation, which is a standard way to solve stochastic programs. (*ii*) We replace the original $\min(\cdot,\cdot)$ operator in constraint (3.3.2) with the two inequality constraints (B.0.14) and (B.0.15). To justify this transformation, we only need to show that at the optimal solution, either (B.0.14) or (B.0.15) will be binding. This condition is equivalent to the argument that in the optimal solution the firm has no incentive to deliberately withhold its production and backorder some demand into the next period, which is obvious since the profit margin is decreasing over time. (*iii*) We replace the original constraint (3.3.5) containing the $(\cdot)^+$ operator with the new linear constraint (B.0.16), which is a common technique. □

**Proof of Proposition 3.3.2:** (i) We prove this result using a simple contradiction argument. When solving stage $n$'s problem, assume that for a certain period $m \ge n + L_b > n + L_f$ (no orders have been committed for period $m$ yet), the optimal solution $B_m^*$ and $F_m^*$ satisfy that $B_m^* = b_m \ge 0$ and $F_m^* = f_m > 0$. Then under the condition that $B^T$ and $F^T$ are not binding, by changing the solution to $\hat{B}_m = b_m + f_m$ and $\hat{F}_m = 0$, we will still be able to satisfy all the constraints while strictly improve the objective function (3.3.1), since $c_b(b_m + f_m) < c_b b_m + c_f f_m$. This contradicts the optimality of $B_m^*$ and $F_m^*$, and hence we must have $F_m^* = 0$. (ii) The number of free decision variables during period $n$, $\Xi_b$ and $\Xi_f$, can be identified based on Figure 3.5 as well as part (i) of Proposition 3.3.2. □

**Proof of Proposition 3.3.3:** From Proposition 3.3.1 we know that the objective function of the execution problem (2) is linear in decisions $\vec{B}$ and $\vec{F}$, and therefore trivially concave in $(\vec{B}, \vec{F}, B^T, F^T)$. Also notice that the constraint set is a convex set. Hence, applying the concavity preservation theorem under maximization (or convexity preservation under minimization) (Heyman and Sobel 1984, p. 525), we know that the value function of the execution problem $J^m(B^T, F^T, \vec{\mu}^m, \vec{\sigma}^m)$ (and thus the objective function of the reservation problem) is concave in $(B^T, F^T)$. To see coerciveness, notice that the objective function value of (3.3.10) goes to negative infinity as $B^T$ or $F^T$ tends to infinity, given that demand is finite. □

**Proof of Proposition A.2.1:** Since $f(x, \zeta)$ is concave in $x$, $g(\cdot)$ is concave and increasing, we know that both $\mathbb{E}_\zeta f(x, \zeta)$ and $\mathbb{E}_\zeta g(f(x, \zeta))$ are concave in $x$. By first order condition and the interchange of integral and differentiation (Cheng 2010), we have that

$$\begin{aligned} \frac{\partial \mathbb{E}_\zeta f(x,\zeta)}{\partial x}|_{x=x^*} &= \mathbb{E}_\zeta f^{(1)}(x^*,\zeta) = \mathbb{E}_\zeta f^{(1)}(x^*,\zeta) \mathrm{I}_{f^{(1)}(x^*,\zeta)\geq 0} \\ &\quad + \mathbb{E}_\zeta f^{(1)}(x^*,\zeta) \mathrm{I}_{f^{(1)}(x^*,\zeta)<0} = 0. \qquad \text{(B.0.17)} \end{aligned}$$

and that $\hat{x}^* \geq x^*$ if and only if $\frac{\partial \mathbb{E}_\zeta g(f(x,\zeta))}{\partial x}|_{x=x^*} = \mathbb{E}_\zeta g'(f(x^*,\zeta)) f^{(1)}(x^*,\zeta) \geq 0$; $\hat{x}^* \leq x^*$ if and only if $\frac{\partial \mathbb{E}_\zeta g(f(x,\zeta))}{\partial x}|_{x=x^*} = \mathbb{E}_\zeta g'(f(x^*,\zeta)) f^{(1)}(x^*,\zeta) \leq 0$ (Here a superscript $(i)$ means the partial derivative with respect to the $i$-th argument; $\mathrm{I}_{(\cdot)}$ represents the indicator function).

Case $(i)$&$(ii)$: Note that the submodularity (supermodularity) of $f(x, \zeta)$ implies that $f^{(1)}(x, \zeta)$ is decreasing (increasing) in $\zeta$. Let $\zeta^*$ be such that $f^{(1)}(x^*, \zeta^*) = 0$, then $f^{(1)}(x^*, \zeta) \geq 0$ implies that $\zeta \leq (\geq) \zeta^*$, which further suggests that $f(x^*, \zeta) \leq f(x^*, \zeta^*)$ (since $f(x, \zeta)$ is increasing (decreasing) in $\zeta$) and $g'(f(x^*, \zeta)) \geq g'(f(x^*, \zeta^*)) \geq 0$ (since $g(\cdot)$ is concave increasing). Similarly, $f^{(1)}(x^*, \zeta) < 0$ implies that $\zeta \geq (\leq) \zeta^*$, which further suggests that $f(x^*, \zeta) \geq f(x^*, \zeta^*)$ (since $f(x, \zeta)$ is increasing (decreasing) in $\zeta$) and $g'(f(x^*, \zeta^*)) \geq g'(f(x^*, \zeta)) \geq 0$. Therefore, we have that

$$\begin{aligned} & \mathbb{E}_\zeta g'(f(x^*,\zeta)) f^{(1)}(x^*,\zeta) \\ = \; & \mathbb{E}_\zeta g'(f(x^*,\zeta)) f^{(1)}(x^*,\zeta) \mathrm{I}_{f^{(1)}(x^*,\zeta)\geq 0} + \mathbb{E}_\zeta g'(f(x^*,\zeta)) f^{(1)}(x^*,\zeta) \mathrm{I}_{f^{(1)}(x^*,\zeta)<0} \\ \geq \; & \mathbb{E}_\zeta g'(f(x^*,\zeta^*)) f^{(1)}(x^*,\zeta) \mathrm{I}_{f^{(1)}(x^*,\zeta)\geq 0} + \mathbb{E}_\zeta g'(f(x^*,\zeta^*)) f^{(1)}(x^*,\zeta) \mathrm{I}_{f^{(1)}(x^*,\zeta)<0} \\ = \; & g'(f(x^*,\zeta^*)) [\mathbb{E}_\zeta f^{(1)}(x^*,\zeta) \mathrm{I}_{f^{(1)}(x^*,\zeta)\geq 0} + \mathbb{E}_\zeta f^{(1)}(x^*,\zeta) \mathrm{I}_{f^{(1)}(x^*,\zeta)<0}] = 0. \qquad \text{(B.0.18)} \end{aligned}$$

This implies that $\hat{x}^* = \arg\max_x \mathbb{E}_\zeta g(f(x,\zeta)) \geq x^*$.

Case $(iii)\&(iv)$: The submodularity (supermodularity) of $f(x,\zeta)$ implies that $f^{(1)}(x,\zeta)$ is decreasing (increasing) in $\zeta$. Let $\zeta^*$ be such that $f^{(1)}(x^*,\zeta^*) = 0$, then $f^{(1)}(x^*,\zeta) \geq 0$ implies that $\zeta \leq (\geq)\zeta^*$, which further suggests that $f(x^*,\zeta) \geq f(x^*,\zeta^*)$ (since $f(x,\zeta)$ is decreasing (increasing) in $\zeta$) and $g'(f(x^*,\zeta^*)) \geq g'(f(x^*,\zeta)) \geq 0$ (since $g(\cdot)$ is concave increasing). Similarly, $f^{(1)}(x^*,\zeta) < 0$ implies that $\zeta \geq (\leq)\zeta^*$, which further suggests that $f(x^*,\zeta) \leq f(x^*,\zeta^*)$ (since $f(x,\zeta)$ is decreasing (increasing) in $\zeta$) and $g'(f(x^*,\zeta)) \geq g'(f(x^*,\zeta^*)) \geq 0$. Therefore, we have that

$$\begin{aligned}
& \mathbb{E}_\zeta g'(f(x^*,\zeta))f^{(1)}(x^*,\zeta) \\
= \; & \mathbb{E}_\zeta g'(f(x^*,\zeta))f^{(1)}(x^*,\zeta)\mathrm{I}_{f^{(1)}(x^*,\zeta)\geq 0} + \mathbb{E}_\zeta g'(f(x^*,\zeta))f^{(1)}(x^*,\zeta)\mathrm{I}_{f^{(1)}(x^*,\zeta)<0} \\
\leq \; & \mathbb{E}_\zeta g'(f(x^*,\zeta^*))f^{(1)}(x^*,\zeta)\mathrm{I}_{f^{(1)}(x^*,\zeta)\geq 0} + \mathbb{E}_\zeta g'(f(x^*,\zeta^*))f^{(1)}(x^*,\zeta)\mathrm{I}_{f^{(1)}(x^*,\zeta)<0} \\
= \; & g'(f(x^*,\zeta^*))[\mathbb{E}_\zeta f^{(1)}(x^*,\zeta)\mathrm{I}_{f^{(1)}(x^*,\zeta)\geq 0} + \mathbb{E}_\zeta f^{(1)}(x^*,\zeta)\mathrm{I}_{f^{(1)}(x^*,\zeta)<0}] \\
= \; & 0. \qquad \text{(B.0.19)}
\end{aligned}$$

This implies that $\hat{x}^* = \arg\max_x \mathbb{E}_\zeta g(f(x,\zeta)) \leq x^*$. □

**Proof of Corollary A.2.2:** We demonstrate below how to derive Case $(vi)$ based on Case $(ii)$, and the other cases should be similar. Note that Case $(vi)$ corresponds to Scenario A in Figure A.1(a). Let $\hat{\zeta}$ be the mode of $f(x^*,\zeta^*)$, $\zeta' \in [\hat{\zeta},\bar{\zeta}]$ be such that $f(x^*,\zeta') = f(x^*,\underline{\zeta}) = \Delta$. Since $g(f)$ is concave increasing for $f < \Delta$ and linear increasing for $f \geq \Delta$, we know $g'(f)$ is positive decreasing for $f < \Delta$ and positive constant for $f \geq \Delta$. Combining this with the fact that $f(x^*,\zeta)$ is less than $\Delta$ and decreasing for $\zeta \geq \zeta'$, we know that $g'(f(x^*,\zeta))$ is weakly increasing in $\zeta$. Now, still let $\zeta^*$ be such that $f^{(1)}(x^*,\zeta^*) = 0$; given the supermodularity of $f(x,\zeta)$, we then know $f^{(1)}(x^*,\zeta) \geq 0$ implies that $\zeta \geq \zeta^*$ and hence $g'(f(x^*,\zeta)) \geq g'(f(x^*,\zeta^*))$, and $f^{(1)}(x^*,\zeta) < 0$ implies that $\zeta < \zeta^*$ and hence $g'(f(x^*,\zeta)) < g'(f(x^*,\zeta^*))$. Following the same logic of Equation B.0.19, we can obtain that $\mathbb{E}_\zeta g'(f(x^*,\zeta))f^{(1)}(x^*,\zeta) \geq 0$, and therefore $\hat{x}^* \geq x^*$. □

**Proof of Proposition 4.4.1:** Denote $\hat{\Pi}_S(\Lambda) = \mathbb{E}\Pi_S(\Lambda) - kVar\Pi_S(\Lambda) = w\Lambda + \mu_s(K-\Lambda) - cK - k(K-\Lambda)^2\sigma_s^2$. The first order derivative is $\frac{\partial\hat{\Pi}_S(\Lambda)}{\partial\Lambda} = (w-\mu_s) + 2k\sigma_s^2(K-\Lambda)$, with a zero given by $\hat{\Lambda} = K - \frac{\mu_s - w}{2k\sigma_s^2}$. The second order derivative is $\frac{\partial^2\hat{\Pi}_S(\Lambda)}{\partial\Lambda^2} = -2k\sigma_s^2 \leq$

0, confirming the concavity of the objective function. Hence, $\hat{\Lambda}$ is the optimal decision if and only if $0 \leq K - \frac{\mu_s - w}{2k\sigma_s^2} < K$, which corresponds to $\mu_s \in (w,\ w + 2k\sigma_s^2 K]$. Therefore the result follows. □

**Proof of Proposition 4.4.2:** Still, we need to check the first and second derivative of the objective function:

$$\frac{\partial \mathbb{E}\Pi_S(\Lambda)}{\partial \Lambda} = (w - \mu_s + \rho\frac{\sigma_s}{\sigma_d}\mu_d)(1 - \Phi_d(\Lambda)) - \rho\frac{\sigma_s}{\sigma_d}\int_{\Lambda}^{\infty} x\phi_d(x)dx, \quad \text{(B.0.20)}$$

$$\begin{aligned}\frac{\partial^2 \mathbb{E}\Pi_S(\Lambda)}{\partial \Lambda^2} &= -(w - \mu_s + \rho\frac{\sigma_s}{\sigma_d}\mu_d)\phi_d(\Lambda) + \rho\frac{\sigma_s}{\sigma_d}\Lambda\phi_d(\Lambda) \\ &= [\rho\frac{\sigma_s}{\sigma_d}(\Lambda - \mu_d) - (w - \mu_s)]\phi_d(\Lambda). \quad \text{(B.0.21)}\end{aligned}$$

Obviously, $\frac{\partial^2 \mathbb{E}\Pi_S(\Lambda)}{\partial \Lambda^2} \leq 0$ if and only if $\rho\frac{\sigma_s}{\sigma_d}(\Lambda - \mu_d) - (w - \mu_s) \leq 0$.

Case (i): If $\rho \geq 0$, $\frac{\partial^2 \mathbb{E}\Pi_S(\Lambda)}{\partial \Lambda^2} \leq 0$ implies $\Lambda \leq \mu_d + \frac{(w-\mu_s)\sigma_d}{\rho\sigma_s}$. Hence $\mathbb{E}\Pi_S(\Lambda)$ is concave-convex on interval $[0, K]$ (purely concave if $K \leq \mu_d + \frac{(w-\mu_s)\sigma_d}{\rho\sigma_s}$). Now, we claim that on the potential convex interval where $\rho\frac{\sigma_s}{\sigma_d}(\Lambda - \mu_d) \geq w - \mu_s$, $\mathbb{E}\Pi_S(\Lambda)$ must be decreasing. To see this, note that

$$\begin{aligned}\frac{\partial \mathbb{E}\Pi_S(\Lambda)}{\partial \Lambda} &= (w - \mu_s + \rho\frac{\sigma_s}{\sigma_d}\mu_d)(1 - \Phi_d(\Lambda)) - \rho\frac{\sigma_s}{\sigma_d}\int_{\Lambda}^{\infty} x\phi_d(x)dx \\ &\leq (w - \mu_s + \rho\frac{\sigma_s}{\sigma_d}\mu_d)(1 - \Phi_d(\Lambda)) - \rho\frac{\sigma_s}{\sigma_d}\int_{\Lambda}^{\infty} \Lambda\phi_d(x)dx \\ &= [w - \mu_s - \rho\frac{\sigma_s}{\sigma_d}(\Lambda - \mu_d)](1 - \Phi_d(\Lambda)) \\ &\leq 0 \quad \text{(B.0.22)}\end{aligned}$$

Also, when $\mu_d$ is large enough so that $\phi_d(0) \approx 0$, we have $\lim_{\Lambda \to 0} \frac{\partial \mathbb{E}\Pi_S(\Lambda)}{\partial \Lambda} = (w - \mu_s + \rho\frac{\sigma_s}{\sigma_d}\mu_d)(1 - \Phi_d(0)) - \rho\frac{\sigma_s}{\sigma_d}\int_0^{\infty} x\phi_d(x)dx \approx w - \mu_s + \rho\frac{\sigma_s}{\sigma_d}\mu_d - \rho\frac{\sigma_s}{\sigma_d}\mu_d = w - \mu_s$.

Therefore, if $w \geq \mu_s$, we know that $\mathbb{E}\Pi_S(\Lambda)$ starts with a concave increasing segment, and hence the maximizer is given by $\min(K, \hat{\Lambda})$, where $\hat{\Lambda}$ is the solution to the first order condition; if $w < \mu_s$, however, we know that $\mathbb{E}\Pi_S(\Lambda)$ is monotonically decreasing on $[0, K]$, and hence 0 is the maximizer.

Case (ii): If $\rho < 0$, $\frac{\partial^2 \mathbb{E}\Pi_S(\Lambda)}{\partial \Lambda^2} \leq 0$ implies $\Lambda \geq \mu_d + \frac{(w-\mu_s)\sigma_d}{\rho\sigma_s}$. Hence $\mathbb{E}\Pi_S(\Lambda)$ is convex-concave on interval $[0, K]$ (purely convex if $K \leq \mu_d + \frac{(w-\mu_s)\sigma_d}{\rho\sigma_s}$). Applying an analysis symmetric to (B.0.22), we can show $\mathbb{E}\Pi_S(\Lambda)$ must be increasing on its concave interval. Therefore, we conclude that $\mathbb{E}\Pi_S(\Lambda)$ is actually quasi-convex and thus an extreme policy is optimal. In particular, if $\mu_s \leq w$, then $\mathbb{E}\Pi_S(\Lambda)$ is monotonically increasing and $\Lambda = K$ is optimal. □

**Proof of Proposition 4.4.3:** Since the objective function $\mathbb{E}\Pi_S(\Lambda)$ is continuous and twice differentiable, we apply Topkis's Theorem to investigate the comparative statics property. Recall that $\frac{\partial \mathbb{E}\Pi_S(\Lambda)}{\partial \Lambda} = (w - \mu_s + \rho\frac{\sigma_s}{\sigma_d}\mu_d)(1 - \Phi_d(\Lambda)) - \rho\frac{\sigma_s}{\sigma_d}\int_\Lambda^\infty x\phi_d(x)dx$; we thus have

$$
\begin{aligned}
\frac{\partial^2 \mathbb{E}\Pi_S(\Lambda)}{\partial\Lambda\partial w} &= 1 - \Phi_d(\Lambda) \geq 0;\\
\frac{\partial^2 \mathbb{E}\Pi_S(\Lambda)}{\partial\Lambda\partial \mu_d} &= \rho\frac{\sigma_s}{\sigma_d}(1 - \Phi_d(\Lambda)) \geq 0;\\
\frac{\partial^2 \mathbb{E}\Pi_S(\Lambda)}{\partial\Lambda\partial \mu_s} &= -(1 - \Phi_d(\Lambda)) \leq 0;\\
\frac{\partial^2 \mathbb{E}\Pi_S(\Lambda)}{\partial\Lambda\partial \sigma_d} &= -\rho\frac{\sigma_s}{\sigma_d^2}\mu_d(1 - \Phi_d(\Lambda)) + \rho\frac{\sigma_s}{\sigma_d^2}\int_\Lambda^\infty x\phi_d(x)dx\\
&\geq -\rho\frac{\sigma_s}{\sigma_d^2}\mu_d(1 - \Phi_d(\Lambda)) + \rho\frac{\sigma_s}{\sigma_d^2}\Lambda\int_\Lambda^\infty \phi_d(x)dx\\
&= (\Lambda - \mu_d)\rho\frac{\sigma_s}{\sigma_d^2}\mu_d(1 - \Phi_d(\Lambda))\\
&\geq 0 \quad (\text{if } \Lambda \geq \mu_d);\\
\frac{\partial^2 \mathbb{E}\Pi_S(\Lambda)}{\partial\Lambda\partial \sigma_s} &= \frac{\rho}{\sigma_d}\mu_d(1 - \Phi_d(\Lambda)) - \frac{\rho}{\sigma_d}\int_\Lambda^\infty x\phi_d(x)dx\\
&\leq \frac{\rho}{\sigma_d}\mu_d(1 - \Phi_d(\Lambda)) - \frac{\rho}{\sigma_d}\Lambda\int_\Lambda^\infty \phi_d(x)dx\\
&= (\mu_d - \Lambda)\frac{\rho}{\sigma_d}(1 - \Phi_d(\Lambda))\\
&\leq 0 \quad (\text{if } \Lambda \geq \mu_d);\\
\frac{\partial^2 \mathbb{E}\Pi_S(\Lambda)}{\partial\Lambda\partial \rho} &= \frac{\sigma_s}{\sigma_d}\mu_d(1 - \Phi_d(\Lambda)) - \frac{\sigma_s}{\sigma_d}\int_\Lambda^\infty x\phi_d(x)dx\\
&\leq \frac{\sigma_s}{\sigma_d}\mu_d(1 - \Phi_d(\Lambda)) - \frac{\sigma_s}{\sigma_d}\Lambda\int_\Lambda^\infty \phi_d(x)dx\\
&= (\mu_d - \Lambda)\frac{\sigma_s}{\sigma_d}(1 - \Phi_d(\Lambda))\\
&\leq 0 \quad (\text{if } \Lambda \geq \mu_d).
\end{aligned}
$$

□

**Proof of Lemma 4.5.2:** We know that $r(z) \geq r(x)$ for $z \in [x, M]$ due to IFR. Therefore,

$$r(x)\int_x^M (1-\Phi(z))dz = \int_x^M \frac{r(x)}{r(z)}\phi(z)dz \leq \int_x^M \phi(z)dz = 1-\Phi(x),$$

which implies that $r(x) \leq \frac{1-\Phi(x)}{\int_x^M (1-\Phi(z))dz}$. □

**Proof of Proposition 4.5.3:** To prove the quasiconvexity of $\Pi_S(\Lambda)$, it is sufficient to verify the following four claims: (i) $\Pi_S(\Lambda)$ is convex for $\Lambda$ close to zero; (ii) If $\Pi_S(\Lambda)$ is concave on an interval, it has to be increasing on this interval; (iii) $\Pi_S(\Lambda)$ is a constant function for a sufficiently large $\Lambda$; (iv) The first order derivative of $\Pi_S(\Lambda)$ is continuous. To start, we first derive the first order derivative and the second order derivative of $\Pi_S(\Lambda)$. Note that $\Pi_S(\Lambda)$ can be explicitly written as

$$\begin{aligned}\Pi_S(\Lambda) &= \mathbb{E}_{D,\epsilon}[(w-c)\min(D,\Lambda)+\theta(a+b(D-\Lambda)^+ + \epsilon - c)(D-\Lambda)^+] \\ &= (w-c)\mathbb{E}\min(D,\Lambda)+\theta(a-c)\mathbb{E}(D-\lambda)^+ + \theta b\int_\Lambda^M (x-\Lambda)^2 d\Phi(x).\end{aligned}$$

The first order derivative is given by

$$\begin{aligned}\Pi_S(\Lambda)' &= (w-c)(1-\Phi(\Lambda))-\theta(a-c)(1-\Phi(\Lambda))-2\theta b(M-\Lambda-\int_\Lambda^M \Phi(x)dx) \\ &= (w-\theta a-\bar{\theta}c)(1-\Phi(\Lambda))-2\theta b\int_\Lambda^M (1-\Phi(x))dx; \qquad \text{(B.0.23)}\end{aligned}$$

and the second order derivative can be further derived as

$$\Pi_S(\Lambda)'' = 2\theta b(1-\Phi(\Lambda))-(w-\theta a-\bar{\theta}c)\phi(\Lambda). \qquad \text{(B.0.24)}$$

Claim (i): For $\Lambda$ close to zero, we have $\phi(\Lambda) = \Phi(\Lambda) = 0$, and $\Pi_S(\Lambda)''|_{\Lambda\to 0} = 2\theta b \geq 0$, which implies that $\Pi_S(\Lambda)$ is convex in this region.

Claim (ii): Assume $\Pi_S(\Lambda)$ is concave on interval $\mathcal{I}$; then we know that $\Pi_S(\Lambda)'' = 2\theta b(1-\Phi(\Lambda))-(w-\theta a-\bar{\theta}c)\phi(\Lambda) \leq 0$ for $\Lambda \in \mathcal{I}$, which implies $(w-\theta a-\bar{\theta}c) \geq$

$\frac{2\theta b(1-\Phi(\Lambda))}{\phi(\Lambda)}$. Hence, we know the first order derivative

$$\begin{aligned}\Pi_S(\Lambda)' &= (w-\theta a-\bar{\theta}c)(1-\Phi(\Lambda))-2\theta b\int_\Lambda^M(1-\Phi(x))dx\\ &\geq \frac{2\theta b(1-\Phi(\Lambda))^2}{\phi(\Lambda)}-2\theta b\int_\Lambda^M(1-\Phi(x))dx\\ &= 2\theta b(1-\Phi(\Lambda))[\frac{1-\Phi(\Lambda)}{\phi(\Lambda)}-\frac{\int_\Lambda^M(1-\Phi(x))dx}{1-\Phi(\Lambda)}]\geq 0,\end{aligned}$$

where the last step is based on Lemma 4.5.2. Therefore, $\Pi_S(\Lambda)$ is also increasing on interval $\mathcal{I}$.

Claim (iii): Obviously, for $\Lambda > M$, we have $\Pi_S(\Lambda) = (w-c)\mu_d$ constant.

Claim (iv): This can be verified by checking the first order derivative expressed in Equation (B.0.23).

The above four points together imply that $\Pi_S(\Lambda)$ is quasiconvex on $[0,\infty]$. □

**Proof of Proposition 4.5.4:** According to Proposition 4.5.3, the quasiconvexity of $\Pi_S(\Lambda)$ implies that the maximum value is attained at the boundary of the domain. Hence, we only need to compare $\Pi_S(0)$ with $\Pi_S(\infty)$, where $\Pi_S(0) = \theta(a-c)\mu_d + \theta b(\mu_d^2+\sigma_d^2)$, and $\Pi_S(\infty) = (w-c)\mu_d$, and let $\Xi = \Pi_S(0) - \Pi_S(\infty)$. □

**Proof of Proposition 4.5.6:** Buyer $i$'s objective function (4.5.6) can be rewritten as

$$\Pi_{Bi}(q_i,\vec{q}_{-i}) \quad = \quad p_m\mu_i - wq_i - (a+b\sum_{j\neq i}\mathbb{E}(D_j-q_j)^+)\mathbb{E}(D_i-q_i)^+ - b\mathbb{E}[(D_i-q_i)^+]^2.$$

The first order derivative with respect to $q_i$ is derived as

$$\frac{\partial\Pi_{Bi}(q_i,\vec{q}_{-i})}{\partial q_i} = (a+b\sum_{j\neq i}\mathbb{E}(D_j-q_j)^+)(1-\Phi_i(q_i)) + 2b\int_{q_i}^{M_i}(1-\Phi_i(x))dx - w; \tag{B.0.25}$$

the second order derivative with respect to $q_i$ is given by

$$\frac{\partial^2 \Pi_{Bi}(q_i, \vec{q}_{-i})}{\partial q_i^2} = -(a + b\sum_{j\neq i}\mathbb{E}(D_j - q_j)^+)\phi_i(q_i) - 2b(1-\Phi_i(q_i)) \leq 0; \quad \text{(B.0.26)}$$

and the cross-partial with respect to $(q_i, q_j)$ is given by

$$\frac{\partial^2 \Pi_{Bi}(q_i, \vec{q}_{-i})}{\partial q_i \partial q_j} = -b(1-\Phi_i(q_i))(1-\Phi_j(q_j)) \leq 0. \quad \text{(B.0.27)}$$

From (B.0.26) we conclude that $\Pi_{Bi}(q_i, \vec{q}_{-i})$ is concave in $q_i$. From (B.0.27) we conclude that $\Pi_{Bi}(q_i, \vec{q}_{-i})$ is submodular in $(q_i, q_j)$ for $j \neq i$. The unconstrained maximizer $\hat{q}_i(\vec{q}_{-i})$ can be implicitly obtained by setting (B.0.25) equal to zero. □

**Proof of Proposition 4.5.8:** First, due to symmetry, the unconstrained optimizer $\hat{q}$ can be obtained through replacing all the $q_j$'s in Equation (4.5.7) with $\hat{q}$, leading to the new F.O.C. given by Equation (4.5.8). Next, to show that the constrained maximizer $q^*$ is given by $\min(\hat{q}, \Lambda)$, we discuss two cases:

Case (i): $\Lambda \geq \hat{q}$. Obviously, $q^* = \hat{q} = \min(\hat{q}, \Lambda)$ since $\hat{q}$ is the unconstrained optimizer;

Case (ii): $\Lambda < \hat{q}$. For any buyer $i$, we claim $q_i^* = \Lambda$. To see this, we know that for all $j \neq i$, $q_j^* \leq \Lambda < \hat{q}$; and Proposition 4.5.6 says $\Pi_{Bi}(q_i, \vec{q}_{-i})$ is submodular in $(q_i, q_j)$, which implies the unconstrained maximizer $\hat{q}_i(\vec{q}_{-i})$ is decreasing in $q_j$ for all $j \neq i$. Combining these two facts, the new unconstrained maximizer $\hat{q}'_i$ for buyer $i$ facing $q_j^* (\leq \Lambda < \hat{q})$ should satisfy $\hat{q}'_i(\vec{q}^*_{-i}) \geq \hat{q}_i(\vec{\hat{q}}_{-i}) = \hat{q}$, which together with concavity implies that the constrained maximizer $q_i^* = \Lambda$. Applying symmetry, we then have for all buyers $q^* = \Lambda = \min(\hat{q}, \Lambda)$. □

**Proof of Proposition 4.5.9:** First let's show the convexity of $\hat{\Pi}_S(\Lambda)$. Equation (4.5.11) can be expanded and rewritten as

$$\begin{aligned}
\hat{\Pi}_S(\Lambda) &= (w-c)N\Lambda + \theta(a-c)\sum_{i=1}^N \mathbb{E}(D_i - \Lambda)^+ + \theta b \mathbb{E}(\sum_{i=1}^N (D_i - \Lambda)^+)^2 \\
&= (w-c)N\Lambda + \theta(a-c)\sum_{i=1}^N \mathbb{E}(D_i - \Lambda)^+ + \theta b \sum_{i=1}^N \sum_{j\neq i} \mathbb{E}(D_i - \Lambda)^+ \mathbb{E}(D_j - \Lambda)^+ \\
&\quad + \theta b \sum_{i=1}^N \mathbb{E}((D_i - \Lambda)^+)^2.
\end{aligned}$$

The first order derivative with respect to $\Lambda$ is given by

$$\begin{aligned}\hat{\Pi}_S(\Lambda)' &= (w-c)N - \theta(a-c)N(1-\Phi(\Lambda)) - 2N(N-1)\theta b(1-\Phi(\Lambda))\mathbb{E}(D-\Lambda)^+ \\ &\quad -2N\theta b\int_{\Lambda}^{M}(1-\Phi(x))dx, \qquad \text{(B.0.28)}\end{aligned}$$

and the second order derivative is given by

$$\begin{aligned}\hat{\Pi}_S(\Lambda)'' &= \theta(a-c)N\phi(\Lambda) + 2N(N-1)\theta b(\phi(\Lambda)\mathbb{E}(D-\Lambda)^+ + (1-\Phi(\Lambda))^2) \\ &\quad +2N\theta b(1-\Phi(\Lambda)). \qquad \text{(B.0.29)}\end{aligned}$$

We can see that when $a \geq c$, $\hat{\Pi}_S(\Lambda)'' \geq 0$ holds, which implies $\hat{\Pi}_S(\Lambda)$ is indeed convex. Considering also the fact that $\Pi_S(\Lambda)$ is constant for $\Lambda \geq \hat{q}$, we can conclude that $\Pi_S(\Lambda)$ is quasiconvex on $[0, \infty)$. □

**Proof of Proposition 4.5.10:** This claim follows directly from the quasiconvexity result. To determine the optimal withholding policy, the supplier only needs to compare the expected profits at the two ends, $\hat{\Pi}_S(0)$ and $\hat{\Pi}_S(\hat{q})$, where $\hat{\Pi}_S(0) = N\theta(Nb\mu_d^2 + (a-c)\mu_d + b\sigma_d^2)$ by plugging zero into Equation (4.5.11). □

**Proof of Proposition 4.5.11:** Notice that Equation (4.5.17) is a standard quadratic function of $\Lambda_S$ with negative initial coefficient. Hence the conclusion directly follows. □

**Proof of Proposition 4.5.12:** The global optimizer $\hat{\Lambda}_S$ can be achieved if and only if

$$\frac{1}{2g}(a + b(D-\Lambda)^+ - c) \leq K - \min(D, \Lambda), \qquad \text{(B.0.30)}$$

If $D < \Lambda$, then Condition (B.0.30) is equivalent to $D \leq K - \frac{a-c}{2g}$, hence for both $D \leq \Lambda \leq K - \frac{a-c}{2g}$ (case $(i)$) and $D \leq K - \frac{a-c}{2g} < \Lambda$ (case $(iv)$), we have that $\Lambda_S^* = \hat{\Lambda}_S = \frac{1}{2g}(a-c)$; and for $K - \frac{a-c}{2g} < D \leq \Lambda$ (case $(v)$), we have $\Lambda_S^* = K - D$.

If $D \geq \Lambda$, then $\frac{1}{2g}(a + b(D-\Lambda) - c) \leq K - \Lambda$ implies that $D \leq \frac{1}{b}(2gK - (2g - b)\Lambda - (a-c))$, which is possible if and only if $\Lambda \leq \frac{1}{b}(2gK - (2g-b)\Lambda - (a-c))$, i.e., $\Lambda \leq K - \frac{a-c}{2g}$. Hence, when $0 \leq \Lambda \leq K - \frac{a-c}{2g}$ holds, $\Lambda_S^* = \hat{\Lambda}_S = \frac{1}{2g}(a + b(D-\Lambda) - c)$ if $\Lambda < D \leq \frac{1}{b}(2gK - (2g-b)\Lambda - (a-c))$ (case $(ii)$), and $\Lambda_S^* = K - \Lambda$ if $D >$

$\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))$ (case $(iii)$). When $K-\frac{a-c}{2g}<\Lambda\leq K$, however, we must have $\frac{1}{2g}(a+b(D-\Lambda)-c)>K-\Lambda$, and thus $\Lambda_S^*=K-\Lambda$ (case $(vi)$). □

**Proof of Proposition 4.5.13:** Since the explicit expression of $\Pi_S^1(\Lambda)$ depends on the range of $\Lambda$, we start by checking the properties of both $\Pi_{S,1}^1(\Lambda)$ (Equation 4.5.21) and $\Pi_{S,2}^1(\Lambda)$ (Equation 4.5.22).
The first order derivative of $\Pi_{S,1}^1(\Lambda)$ with respect to $\Lambda$ is derived as

$$\begin{aligned}
&\Pi_{S,1}^1(\Lambda)' \\
=\;& (w-c)(1-\Phi(\Lambda))-h+\frac{(a-c)^2}{4g}\phi(\Lambda) \\
&+\int_{\Lambda}^{\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))}-\frac{b}{2g}(a+b(x-\Lambda)-c)\phi(x)dx \\
&+(1-\frac{2g}{b})g(k-\Lambda)^2\phi(\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c)))-\frac{(a-c)^2}{4g}\phi(\Lambda) \\
&+\int_{\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))}^{M}[2(b-g)\Lambda+2gK-b(x+K)-(a-c)]\phi(x)dx \\
&-(1-\frac{2g}{b})g(k-\Lambda)^2\phi(\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c)))-\frac{(a-c)^2}{4g}\phi(\Lambda) \\
=\;& (w-c)(1-\Phi(\Lambda))-h-\int_{\Lambda}^{\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))}\frac{b}{2g}(a+b(x-\Lambda)-c)\phi(x)dx \\
&+\int_{\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))}^{M}[2(b-g)\Lambda+2gK-b(x+K)-(a-c)]\phi(x)dx.
\end{aligned} \tag{B.0.31}$$

The second order derivative of $\Pi_{S,1}^1$ with respect to $\Lambda$ is given by

$$\begin{aligned}
&\Pi_{S,1}^1(\Lambda)'' \\
=\;& -(w-c)\phi(\Lambda)+\int_{\Lambda}^{\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))}\frac{b^2}{2g}dx \\
&-(1-\frac{2g}{b})b(K-\Lambda)\phi(\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c)))+\frac{b}{2g}(a-c)\phi(\Lambda) \\
&+\int_{\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))}^{M}2(b-g)\phi(x)dx
\end{aligned}$$

$$
\begin{aligned}
&+(1-\frac{2g}{b})b(K-\Lambda)\phi(\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c)))+\frac{b}{2g}(a-c)\phi(\Lambda)\\
= \; &(\frac{b}{2g}(a-c)-(w-c))\phi(\Lambda)-2(g-b)[1-\Phi(\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c)))]\\
&+\frac{b^2}{2g}[\Phi(\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c)))-\Phi(\Lambda)].
\end{aligned}
\tag{B.0.32}
$$

Similarly, the first order derivative of $\Pi^1_{S,2}(\Lambda)$ with respect to $\Lambda$ is derived as

$$
\begin{aligned}
\Pi^1_{S,2}(\Lambda)' \;=\; &(w-c)(1-\Phi(\Lambda))-h+\int_{\Lambda}^{M}[2(b-g)\Lambda\\
&+2gK-b(x+K)-(a-c)]\phi(x)dx.
\end{aligned}
\tag{B.0.33}
$$

The second order derivative of $\Pi^1_{S,2}(\Lambda)$ with respect to $\Lambda$ is given by

$$
\Pi^1_{S,2}(\Lambda)'' \;=\; -[(2g-b)(K-\Lambda)+(w-a)]\phi(\Lambda)-2(g-b)(1-\Phi(\Lambda)). \tag{B.0.34}
$$

Since when $\Lambda = K-\frac{a-c}{2g}$ we have $\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c)) = K-\frac{a-c}{2g} = \Lambda$, it can be shown that $\Pi^1_{S,1}(K-\frac{a-c}{2g}) = \Pi^1_{S,2}(K-\frac{a-c}{2g})$ and that $\Pi^1_{S,1}(\Lambda)'|_{\Lambda=K-\frac{a-c}{2g}} = \Pi^1_{S,2}(\Lambda)'|_{\Lambda=K-\frac{a-c}{2g}}$. Therefore, we know that function $\Pi^1_S(\Lambda)$ is continuous and smooth on its domain.

Next, we investigate how the signs of $\Pi^1_{S,1}(\Lambda)''$ and $\Pi^1_{S,2}(\Lambda)''$ change with the value of $\Lambda$. Since $b \le g$ and $a \le w$, the first term of $\Pi^1_{S,1}(\Lambda)''$, i.e., $(\frac{b}{2g}(a-c)-(w-c))\phi(\Lambda)$, is always negative. The remaining terms of $\Pi^1_{S,1}(\Lambda)''$, i.e., $\Gamma_{S,1}(\Lambda) = -2(g-b)[1-\Phi(\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c)))]+\frac{b^2}{2g}[\Phi(\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c)))-\Phi(\Lambda)]$, are decreasing in $\Lambda$ given the fact that $g \ge b$. In particular, note that $\Gamma_{S,1}(K-\frac{a-c}{2g}) = -2(g-b)\Phi(K-\frac{a-c}{2g}) \le 0$.

**Assumption B.0.3.** *The demand pdf $\phi(\cdot)$ is unimodal and its mode $m_D$ satisfies $\Gamma_{S,1}(m_D) \le 0$.*

One sufficient condition that leads to the above assumption is $K-\frac{a-c}{2g} \le m_D$, which, according to our numerical analysis later on, is a very reasonable one. Now, we define $\dot{\Lambda} = \min(m_D, K-\frac{a-c}{2g})$, and, since $\phi(\Lambda)$ is increasing on interval $[0,\dot{\Lambda})$, we can safely argue that $\Pi^1_{S,1}(\Lambda)''$ is monotonically decreasing to negative on $[0, \dot{\Lambda})$ and that it remains negative on interval $[\dot{\Lambda}, K - \frac{a-c}{2g}]$.

When $\Lambda \in (K - \frac{a-c}{2g}, K]$, we switch to $\Pi^1_{S,2}(\Lambda)''$. Note that $\Pi^1_{S,2}(\Lambda)'' \leq 0$ holds trivially when $b \leq g$ and $a \leq w$.

Therefore, we can now conclude the supplier's expected profit function $\Pi^1_S(\Lambda)$ is convex-concave in $\Lambda$ on interval $[0, K]$ with at most one interior maximizer $\hat{\Lambda}$, which would be the unique solution to either $\Pi^1_{S,1}(\Lambda)' = 0$, $(\Lambda \in [0, K - \frac{a-c}{2g}])$ or $\Pi^1_{S,2}(\Lambda)' = 0$, $(\Lambda \in (K - \frac{a-c}{2g}, K])$. □

**Proof of Proposition 4.5.15:** We first demonstrate the monotonicity of the potential interior maximizer $\hat{\Lambda}$. Since the objective function is continuous and twice differentiable, we apply Topkis's Theorem to investigate the comparative statics. Let $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ denote the two zeros of $\Pi^1_{S,1}(\Lambda)'$ and $\Pi^1_{S,2}(\Lambda)'$, respectively; that is, $\Pi^1_{S,1}(\hat{\Lambda}_1)' = 0$ and $\Pi^1_{S,2}(\hat{\Lambda}_2)' = 0$.

$$
\begin{aligned}
\frac{\partial^2 \Pi^1_{S,1}}{\partial \Lambda \partial a} &= -\int_{\Lambda}^{\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))} \frac{b}{2g}\phi(x)dx - \int_{\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))}^{M} \phi(x)dx \leq 0; \\
\frac{\partial^2 \Pi^1_{S,2}}{\partial \Lambda \partial a} &= -\int_{\Lambda}^{M} \phi(x)dx = -(1-\Phi(\Lambda)) \leq 0; \\
\frac{\partial^2 \Pi^1_{S,1}}{\partial \Lambda \partial b} &= -\int_{\Lambda}^{\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))} \frac{a+2b(x-\Lambda)-c}{2g}\phi(x)dx \\
&\quad -\int_{\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))}^{M} (x+K-2\Lambda)\phi(x)dx \leq 0; \\
\frac{\partial^2 \Pi^1_{S,2}}{\partial \Lambda \partial b} &= -\int_{\Lambda}^{M} (x+K-2\Lambda)\phi(x)dx \leq 0; \\
\frac{\partial^2 \Pi^1_{S,1}}{\partial \Lambda \partial g} &= \int_{\Lambda}^{\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))} \frac{b}{2g^2}(a+b(x-\Lambda)-c)\phi(x)dx \\
&\quad +\int_{\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))}^{M} 2(K-\Lambda)\phi(x)dx \geq 0; \\
\frac{\partial^2 \Pi^1_{S,2}}{\partial \Lambda \partial g} &= \int_{\Lambda}^{M} (2K-2\Lambda)\phi(x)dx = 2(K-\Lambda)(1-\Phi(\Lambda)) \geq 0; \\
\frac{\partial^2 \Pi^1_{S,1}}{\partial \Lambda \partial w} &= \frac{\partial^2 \Pi^1_{S,2}}{\partial \Lambda \partial w} = (1-\Phi(\Lambda)) \geq 0;
\end{aligned}
$$

$$\begin{aligned}\frac{\partial^2 \Pi_{S,1}^1}{\partial \Lambda \partial c} &= -(1-\Phi(\Lambda)) + \frac{b}{2g}\int_{\Lambda}^{\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))} \phi(x)dx \\ &\quad + \int_{\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))}^{M} \phi(x)dx \\ &= -(1-\frac{b}{2g})[\Phi(\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))) - \Phi(\Lambda)] \leq 0; \\ \frac{\partial^2 \Pi_{S,2}^1}{\partial \Lambda \partial c} &= -(1-\Phi(\Lambda)) + 1 - \Phi(\Lambda) = 0; \\ \frac{\partial^2 \Pi_{S,1}^1}{\partial \Lambda \partial h} &= \frac{\partial^2 \Pi_{S,2}^1}{\partial \Lambda \partial h} = -1 \leq 0.\end{aligned}$$

We have shown that both $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ are decreasing in $a$, $b$, $c$, $h$ and increasing in $g$, $w$; hence, $\hat{\Lambda}$ will naturally have the same property.

The monotonicity of profit values $\Pi_S^1(\hat{\Lambda})$, $\Pi_S^1(0)$, and $\Pi_S^1(K)$ is more straightforward: they are in general increasing in parameters on the revenue side ($a$, $b$, $w$, etc.) and decreasing in parameters on the cost side ($g$, $c$, $h$, etc.), with the exception that $\Pi_S^1(0)$ is irrelevant with $w$ and $h$ (Equation 4.5.24), and $\Pi_S^1(K)$ is irrelevant with $a$, $b$, and $g$ (Equation 4.5.25). □

**Proof of Proposition 4.5.16:** Facing a deterministic demand $D$, the supplier's first stage quantity decision $\Lambda$ will not exceed $D$. Applying Proposition 4.5.11, the supplier's optimal spot market quantity is given by $\Lambda_S^* = \min(\frac{a+b(D-\Lambda)-c}{2g}, K-\Lambda)$.

When $\frac{a+b(D-\Lambda)-c}{2g} \leq K-\Lambda$, i.e., $\Lambda \leq \frac{2gK-(a+bD-c)}{2g-b}$, we need to further compare $D$ with $\frac{2gK-(a+bD-c)}{2g-b}$: $D \leq \frac{2gK-(a+bD-c)}{2g-b}$ if and only if $D \leq K - \frac{a-c}{2g}$. Hence, when $D \leq K - \frac{a-c}{2g}$, we have $\Lambda \leq D \leq \frac{2gK-(a+bD-c)}{2g-b}$, and $\Lambda_S^* = \frac{a+b(D-\Lambda)-c}{2g}$; when $D > K - \frac{a-c}{2g}$, however, we have $\Lambda_S^* = \frac{a+b(D-\Lambda)-c}{2g}$ if $\Lambda \leq \frac{2gK-(a+bD-c)}{2g-b}$, and $\Lambda_S^* = K-\Lambda$ if $\frac{2gK-(a+bD-c)}{2g-b} < \Lambda \leq D$.

Now, let's look at the supplier's first stage problem. If $D \leq K - \frac{a-c}{2g}$, then the supplier is solving the following maximization in the first stage:

$$\pi_S^1 = \max_{0\leq\Lambda\leq D} \Pi_S^1(\Lambda) = \max_{0\leq\Lambda\leq D} (w-c-h)\Lambda + \frac{(a+b(D-\Lambda)-c)^2}{4g}.$$

Since $[S.O.C.] = \frac{\partial^2 \Pi_S^1(\Lambda)}{\partial \Lambda^2} = \frac{b^2}{2g} \geq 0$, we know $\Pi_S^1$ is convex and hence the optimal solution is of an extreme type. Specifically, we compare $\Pi_S^1(0) = \frac{(a+bD-c)^2}{4g}$ with

$\Pi^1_S(D) = (w-c-h)D + \frac{(a-c)^2}{4g}$ and obtain that $\Pi_S(0) \leq \Pi_S(D)$ if and only if $D \leq \frac{4g(w-c-h)-2b(a-c)}{b^2}$. Therefore, when $D \leq \min(\frac{4g(w-c-h)-2b(a-c)}{b^2}, K - \frac{a-c}{2g})$, we have $\Lambda^* = D$ and $\Lambda^*_S = \frac{a+b(D-D)-c}{2g} = \frac{a-c}{2g}$ (case (i)); when $\frac{4g(w-c-h)-2b(a-c)}{b^2} < D \leq K - \frac{a-c}{2g}$, we have $\Lambda^* = 0$ and $\Lambda^*_S = \frac{a+bD-c}{2g}$ (case (ii)).

If $D > K - \frac{a-c}{2g}$, however, the supplier is solving the following optimization:

$$\pi^1_S = \max\{\max_{0\leq\Lambda\leq\frac{2gK-(a+bD-c)}{2g-b}} \Pi^1_{S,1}(\Lambda), \max_{\frac{2gK-(a+bD-c)}{2g-b}<\Lambda\leq D} \Pi^1_{S,2}(\Lambda)\},$$

where $\Pi^1_{S,1}(\Lambda) = (w-c-h)\Lambda + \frac{(a+b(D-\Lambda)-c)^2}{4g}$ and $\Pi^1_{S,2}(\Lambda) = (w-c-h)\Lambda + (a + b(D-\Lambda) - g(K-\Lambda) - c)(K-\Lambda)$.

We already showed previously that $\Pi^1_{S,1}(\Lambda)$ is convex. Now, let's check the property of $\Pi^1_{S,2}(\Lambda)$:

$$[F.O.C.] \quad \frac{\partial \Pi^1_{S,2}(\Lambda)}{\partial \Lambda} = (w-c-h) + (2g-b)K - (a+bD-c) - 2(g-b)\Lambda;$$
$$[S.O.C.] \quad \frac{\partial^2 \Pi^1_{S,2}(\Lambda)}{\partial \Lambda^2} = -2(g-b) \leq 0.$$

Therefore, we can see that $\Pi^1_{S,2}(\Lambda)$ is actually concave under the condition that $g \geq b$. By setting [F.O.C.]$= 0$, we obtain the interior maximizer $\hat{\Lambda} = \frac{(w-c-h)+(2g-b)K-(a+bD-c)}{2(g-b)}$, which can be achieved if and only if $\hat{\Lambda} \in [\frac{2gK-(a+bD-c)}{2g-b}, D]$, which further translates into a condition that

$$K - \frac{a+h-w}{2g-b} \leq D \leq K - \frac{a-c}{b} + \frac{(2g-b)(w-c-h)}{b^2}.$$

Hence, based on the convex-concave structure of the objective function, if $\max(K - \frac{a-c}{2g}, K - \frac{a+h-w}{2g-b}) \leq D \leq K - \frac{a-c}{b} + \frac{(2g-b)(w-c-h)}{b^2}$, which means the interior maximizer is attainable, we only need to compare $\Pi^1_{S,1}(0) = \frac{(a+bD-c)^2}{4g}$ with $\Pi^1_{S,2}(\hat{\Lambda}) = (w-c-h)\hat{\Lambda} + (a + b(D-\hat{\Lambda}) - g(K-\hat{\Lambda}) - c)(K-\hat{\Lambda})$. If $\Pi^1_{S,1}(\hat{\Lambda}) \geq \Pi^1_{S,1}(0)$, then $\Lambda^* = \hat{\Lambda}$ and $\Lambda^*_S = K - \hat{\Lambda}$; if else, then $\Lambda^* = 0$ and $\Lambda^*_S = \frac{a+bD-c}{2g}$ (case (iii)).

Otherwise if $\hat{\Lambda}$ is not attainable, we only need to compare $\Pi^1_{S,1}(0) = \frac{(a+bD-c)^2}{4g}$ with $\Pi^1_{S,2}(D) = (w-c-h)D + (a - g(K-D) - c)(K-D)$. If $\Pi^1_{S,1}(D) \geq \Pi^1_{S,1}(0)$, then $\Lambda^* = D$ and $\Lambda^*_S = K - D$; if else, then $\Lambda^* = 0$ and $\Lambda^*_S = \frac{a+bD-c}{2g}$ (case (iv)). □

**Proof of Proposition 4.5.17:** Since this is a stochastic setting, we base our analysis on the general discussion in Section 4.5.2. Before start, we make an observation about the integral divider $\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))$ in Equation (4.5.21):

$$\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))\begin{cases}\in[0,D_L), & \text{for } \Lambda>\gamma_2\\ \in[D_L,D_H), & \text{for } \Lambda\in(\gamma_1,\gamma_2]\\ \in[D_H,\frac{1}{b}(2gK-(a-c))], & \text{for } \Lambda\leq\gamma_1\end{cases} \tag{B.0.35}$$

First, we consider $\gamma_1 \leq D_L$. For $\Lambda \leq \gamma_1 (< K-\frac{a-c}{2g})$, we know $\Lambda \leq D_L < D_H \leq \frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))$ from Equation (B.0.35); and hence based on Equation (4.5.21), the supplier's first stage expected profit is

$$\begin{aligned}\Pi_S^1(\Lambda) &= (w-c-h)\Lambda+\frac{1}{4g}[p(a+b(D_H-\Lambda)-c)^2+\bar{p}(a+b(D_L-\Lambda)-c)^2]\\ &= (w-c-h)\Lambda+\frac{b\mu_d}{2g}(a-c-b\Lambda)+\frac{(a-c-b\Lambda)^2}{4g}+\frac{b^2}{4g}(pD_H^2+\bar{p}D_L^2).\end{aligned} \tag{B.0.36}$$

Since $\frac{\partial \Pi_S^1(\Lambda)}{\partial\Lambda}=(w-c-h)-\frac{b^2\mu_d}{2g}-\frac{b}{2g}(a-c-b\Lambda)$ and $\frac{\partial^2\Pi_S^1(\Lambda)}{\partial\Lambda^2}=\frac{b^2}{2g}\geq 0$, we know $\Pi_S^1(\Lambda)$ is convex on $[0,\gamma_1]$.

Now for $\gamma_1<\Lambda\leq D_L(\leq K-\frac{a-c}{2g}\leq\gamma_2)$, we have $\Lambda\leq D_L\leq\frac{1}{b}(2gK-(2g-b)\Lambda-(a-c))<D_H$, and hence

$$\begin{aligned}\Pi_S^1(\Lambda) &= (w-c-h)\Lambda+\frac{\bar{p}}{4g}(a+b(D_L-\Lambda)-c)^2\\ &\quad +p(K-\Lambda)[a+b(D_H-\Lambda)-g(K-\Lambda)-c].\end{aligned} \tag{B.0.37}$$

$$[F.O.C.]\quad \frac{\partial\Pi_S^1(\Lambda)}{\partial\Lambda} = (w-c-h)-p(a+bD_H-c-(2g-b)K) -\bar{p}\frac{b}{2g}(a+bD_L-c)+[\bar{p}\frac{b^2}{2g}-2p(g-b)]\Lambda; \tag{B.0.38}$$

$$[\text{S.O.C.}]\quad \frac{\partial^2\Pi_S^1(\Lambda)}{\partial\Lambda^2} = -2p(g-b)+(1-p)\frac{b^2}{2g}. \tag{B.0.39}$$

We have $\frac{\partial^2\Pi_S^1(\Lambda)}{\partial\Lambda^2}\leq 0$, i.e., $\Pi_S^1(\Lambda)$ being concave on $[\gamma_1,D_L]$, if and only if $p\geq$

$(\frac{b}{2g-b})^2$; and by setting $\frac{\partial \Pi_S^1(\Lambda)}{\partial \Lambda} = 0$, we have the unconstrained optimizer $\hat{\Lambda}_2 = \frac{2g[(w-c-h)-p(a+bD_H-c-(2g-b)K)]-\bar{p}b(a+bD_L-c)}{4pg(g-b)-\bar{p}b^2}$, which can only be attained if it is within $[\gamma_1, D_L]$. If $p \leq (\frac{b}{2g-b})^2$, however, we know $\Pi_S^1(\Lambda)$ continues to be convex on $[\gamma_1, D_L]$.

For $D_L < \Lambda \leq K - \frac{a-c}{2g} (\leq \gamma_2)$, we have $D_L \leq \frac{1}{b}(2gK - (2g-b)\Lambda - (a-c)) < D_H$; this together with the case in which $(D_L \leq) K - \frac{a-c}{2g} \leq \Lambda \leq D_H$ both lead to the following expected profit function:

$$\begin{aligned} \Pi_S^1(\Lambda) &= (w-c)(p\Lambda + \bar{p}D_L) - h\Lambda + \frac{\bar{p}}{4g}(a-c)^2 \\ &\quad + p(K-\Lambda)[a + b(D_H - \Lambda) - g(K-\Lambda) - c] \end{aligned} \tag{B.0.40}$$

$$[F.O.C.] \quad \frac{\partial \Pi_S^1(\Lambda)}{\partial \Lambda} = p[w - a - b(D_H - \Lambda) + (2g-b)(K-\Lambda)] - h; \tag{B.0.41}$$

$$[\text{S.O.C.}] \quad \frac{\partial^2 \Pi_S^1(\Lambda)}{\partial \Lambda^2} = -2p(g-b) \leq 0. \tag{B.0.42}$$

Hence, we know that $\Pi_S^1(\Lambda)$ is concave on interval $(D_L, D_H]$; by setting $\frac{\partial \Pi_S^1(\Lambda)}{\partial \Lambda} = 0$, we have the unconstrained optimizer $\hat{\Lambda}_1 = \frac{p(w-a-bD_H+(2g-b)K)-h}{2p(g-b)}$, which can be achieved if it is within $(D_L, D_H]$.

One can further verify that $\Pi_S^1(\Lambda)$ is smooth at every point but $\Lambda = D_L$, where the left derivative is bigger than the right derivative by $\bar{p}(w - c - \frac{b}{2g}(a-c)) > 0$ and thus a spike might exist. Therefore, Case (i) in Proposition 4.5.17 follows based on the above results. We summarize the supplier's expected profit function here for the reference in the proposition.

$$\Pi_S^1(\Lambda) = \begin{cases} (w-c-h)\Lambda + \frac{b\mu_d}{2g}(a-c-b\Lambda) + \frac{(a-c-b\Lambda)^2}{4g} + \frac{b^2}{4g}(pD_H^2 + \bar{p}D_L^2), \\ \qquad \text{for } \Lambda \in [0, \gamma_1] \\ (w-c-h)\Lambda + \frac{\bar{p}}{4g}(a + b(D_L - \Lambda) - c)^2 \\ \qquad + p(K-\Lambda)[a + b(D_H - \Lambda) - g(K-\Lambda) - c], \quad \text{for } \Lambda \in (\gamma_1, D_L] \\ (w-c)(p\Lambda + \bar{p}D_L) - h\Lambda + \frac{\bar{p}}{4g}(a-c)^2 \\ \qquad + p(K-\Lambda)[a + b(D_H - \Lambda) - g(K-\Lambda) - c], \quad \text{for } \Lambda \in (D_L, D_H] \end{cases} \tag{B.0.43}$$

One can prove the case of $\gamma_1 > D_L$ following a very similar process. We also provide the supplier's expected profit function for this case below as a reference.

$$\Pi_S^1(\Lambda) = \begin{cases} (w-c-h)\Lambda + \frac{b\mu_d}{2g}(a-c-b\Lambda) + \frac{(a-c-b\Lambda)^2}{4g} + \frac{b^2}{4g}(pD_H^2 + \bar{p}D_L^2), \\ \hfill \text{for } \Lambda \in [0, D_L] \\ (w-c)(p\Lambda + \bar{p}D_L) - h\Lambda + \frac{\bar{p}}{4g}(a-c)^2 + \frac{p}{4g}(a+b(D_H-\Lambda)-c)^2, \\ \hfill \text{for } \Lambda \in (D_L, \gamma_1] \\ (w-c)(p\Lambda + \bar{p}D_L) - h\Lambda + \frac{\bar{p}}{4g}(a-c)^2 \\ \quad +p(K-\Lambda)[a+b(D_H-\Lambda)-g(K-\Lambda)-c], \quad \text{for } \Lambda \in (\gamma_1, D_H] \end{cases} \tag{B.0.44}$$

□

# Appendix C

# DMEP Algorithm Flow Chart

**Notations**

BaseTotal: the total reservation quantity with the base mode;

FlexTotal: the total reservation quantity with the flexible mode;

BaseOrders: order quantities from the base mode for different periods under different monte carlo iterations;

FlexOrders: order quantities from the flexible mode for different periods under different monte carlo iterations;

ActSales: actual demand satisfied in each period;

IncmDmd: incoming demand in each period;

CumuDmd: cumulated demand (incoming demand plus previous backorder) in each period;

RemnDmd: remaining unsatisfied demand at the end of each period;

CumuCap: cumulated capacity position at the beginning of each period;

Profit: expected profit for each period;

ProfitOpt: expected total profit across the planning horizon;

Margin: unit profit margin (before the equipment cost is taken out) of the product;

BaseResvPrc: unit reservation price for the base mode;

FlexResvPrc: unit reservation price for the flexible mode;

BaseExePrc: unit execution price for the base mode;

FlexExePrc: unit execution price for the flexible mode;

HoldPrc: unit holding cost of the equipment;

PenltPrc: unit penalty cost for the unmet demand after the entire planning horizon ends;
SerLev: service level target.

**1. The Input/Output Module**

This module is the user-interface where the capacity manager feeds the values of the basic parameters to the heuristic model, the optimization package (here we use CVX) is called and the optimal decision (reservation levels and order quantities) is outputted. Figure C.1 demonstrates the algorithm of the I/O Module.

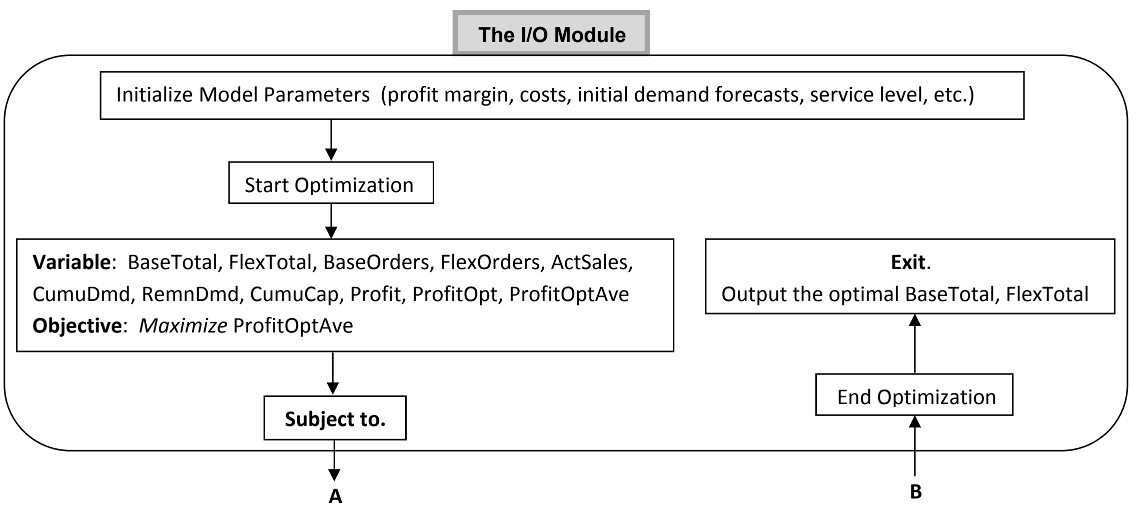



Figure C.1: Flow Diagram of the I/O Module

**Step I.1.** Initialize all the parameters: the contract costs (unit reservation price, unit execution price), the capacity costs (unit holding cost, additional penalty cost), the profit margin, the potential constraints (service level constraint and ramp-up constraint), the initial demand forecasts for all future periods and the belief about how these forecasts could evolve, and for each period the latest demand forecasts as well as the realized demand information. Go to Step I.2.
**Step I.2.** Start the optimization by defining all the (intermediate) decision variables: total reservation quantities for base and flexible modes, base orders and flexible orders for every period under each potential monte carlo iteration, actual sales per

period, cumulated demand per period, remaining unmet demand per period, profits corresponding to each selling period and also the entire selling horizon; and ascertaining the objective function: to maximize the average expected total profit across the six-period selling horizon. Go to Step R.1. in the reservation module.

**Step I.3.** Once quit from the reservation module, end the optimization and output the optimal reservation levels and order quantities for base and flexible modes. Also report the maximum expected total profit across the selling horizon.

## 2. The Reservation Module

Figure C.2 below demonstrates the algorithm of the Reservation Module.

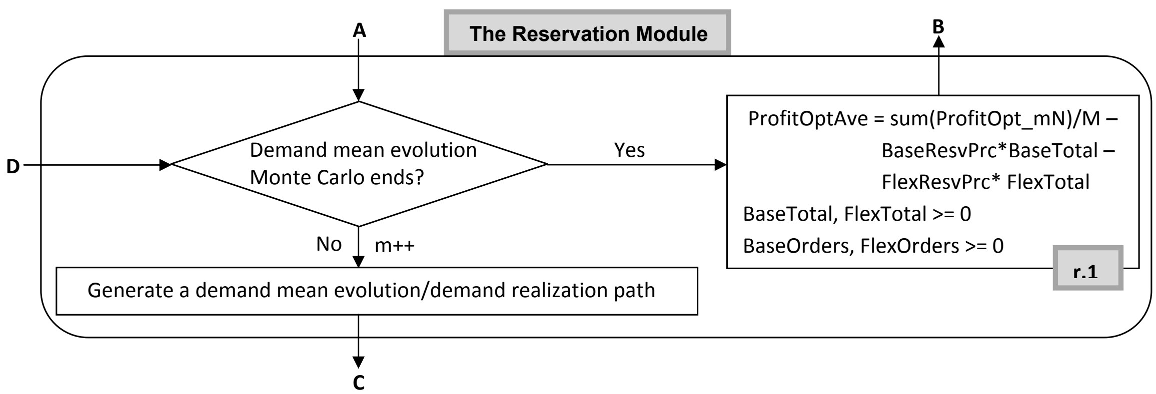



Figure C.2: Flow Diagram of the Reservation Module

**Step R.1.** If the Monte Carlo iteration on demand mean evolution ends, calculate the average expected total profit by averaging the total profits for different mean evolution paths and subtracting the reservation costs for the total base and flexible quantities reserved (refer to equations and constraints in box r.1), then quit the reservation module and go to Step I.3; otherwise go to Step R.2.

**Step R.2.** Generate a demand mean evolution (including demand realization for planning periods in the selling horizon) path and go to Step E.1.

## 3. The Execution Module

Figure C.3 below demonstrates the algorithm of the Execution Module.

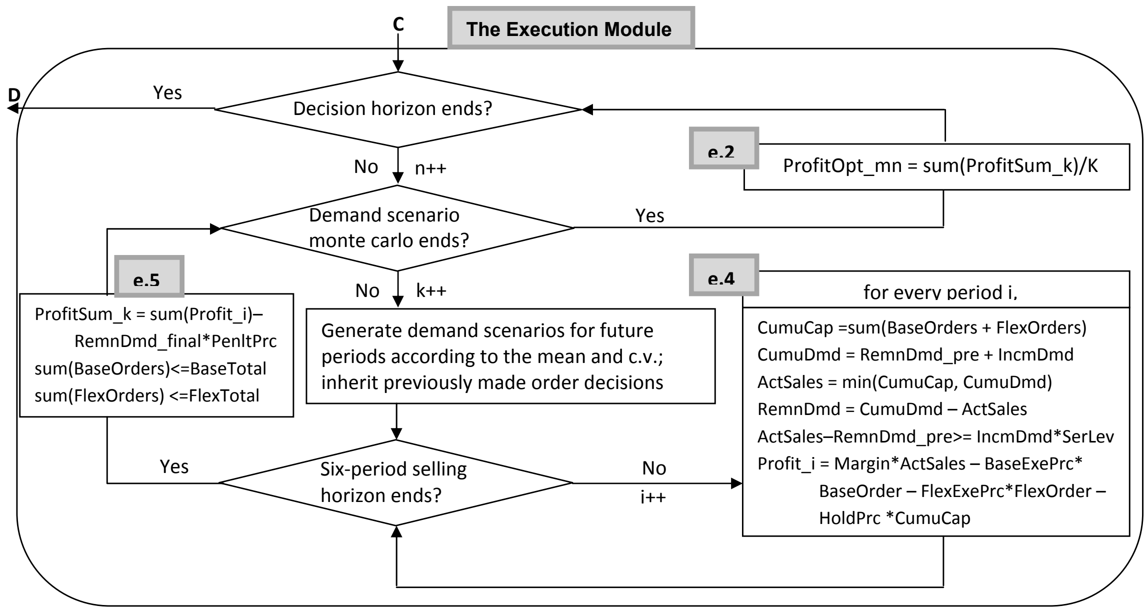



Figure C.3: Flow Diagram of the Execution Module

**Step E.1.** If the decision horizon ends, quit the execution module and go to Step R.1.; otherwise, move to the next planning period and go to Step E.2.

**Step E.2.** If the Monte Carlo iteration on demand scenarios ends, calculate the expected total profit (when standing at a particular planning period) by averaging the profits obtained during each Monte Carlo iteration (refer to equation in box e.2.), then return to Step E.1.; otherwise, go to Step E.3.

**Step E.3.** Generate demand scenarios for future periods according to the demand mean values obtained in Step R.2.; recall the previously made order decisions up to the current planning period. Go to Step E.4.

**Step E.4.** Calculate the profit for each demand period based on/subject to the balance equations and constraints in box e.4.; once done, calculate the total profit of the entire selling horizon based on/subject to the equations and constraints in box e.5. Return to Step E.2.

# Bibliography


[1] B. Allaz. Oligopoly, uncertainty and strategic forward transactions. *International Journal of Industrial Organization*, 10:297–308, 1992.

[2] A. Angelus and E.L. Porteus. Simultaneous capacity and production management of short-life-cycle, produce-to-stock goods under stochastic demand. *Management Science*, 48(3):399–413, 2002.

[3] D. Armbrust. Now is the time. *SEMATECH Board Meeting.*

[4] P. Ausick. Rio tinto boosts iron ore production but faces speed bumps. *Wall Street Journal*, April 15, 2010.

[5] K. Baker. An experimental study of the effectiveness of rolling schedules in production planning. *Decision Science*, 8:19–27, 1977.

[6] K. Baker and D. Peterson. An analytical framework for evaluating the rolling schedules. *Management Science*, 25:341–351, 1979.

[7] Y. Bassok and R. Anupindi. Analysis of supply contracts with total minimum commitment. *IIE Transactions*, 29(5):373–381, 1997.

[8] Y. Bassok, A. Bixby, and R. Srinivasan. Design of component-supply contract with commitment-revision flexibility. *IBM Journal of Research and Development*, 41(6), 1999.

[9] C. Billington and B. Johnson. A real options perspective on supply management in high technology. *Journal of Applied Corporate Finance*, Summer:20–28, 2002.

[10] G.R. Bitran and T.Y. Leong. Deterministic approximations to co-production problems with service constraints and random yields. *Management Science*, 38(5):724–742, 1992.

[11] G.R. Bitran and H.H. Yanasse. Deterministic approximations to stochastic production problems. *Operations Research*, 32(5):999–1018, 1984.

[12] J. Blas. Iron ore pricing emerges from stone age. *Financial Times*, October 26, 2009.

[13] S. Burns. Alumina pricing could follow iron ore to a quarterly or spot contract. *MetalMiner*, April 20, 2010.

[14] G. P. Cachon. Supply chain coordination with contracts. In A. G. de Kok and S. C. Graves, editors, *Handbooks in Operations Research and Management Science - Supply Chain Management: Design, Coordination and Operation*, chapter 6, pages 229–340. Elsevier, Amsterdam, The Netherlands, 2003.

[15] X. Chao, H. Chen, and S. Zheng. Dynamic capacity expansion for a service firm with capacity deterioration and supply uncertainty. *Operations Research*, 57(1):82–93, 2009.

[16] K.Y. Chen, M. Kaya, and ´O. ´Ozer. Dual sales channel management with service competition. *Manufacturing & Service Operations Management*, 10(4):654–675, 2008.

[17] L. Chen and H.L. Lee. Information sharing and order variability control under a generalized demand model. *Management Science*, 55(5):781–797, 2009.

[18] S. Cheng. Differentiation under the integral sign with weak derivatives. `http://www.gold-saucer.org/math/`, 2010.

[19] W.K. Chiang, D. Chhajed, and J.D. Hess. Direct marketing, indirect profits: a strategic analysis of dual-channel supply-chain design. *Management Science*, 49(1):269–286, 2003.

[20] M. Cohen, T.H. Ho, Z.J. Ren, and C. Terwiesch. Measuring imputed cost in the semiconductor equipment supply chain. *Management Science*, 49(12):1653–1670, 2003.

[21] K.H. Daniel. A delivery-lag inventory model with emergency. In H.E. Scarf, D.M. Gilford, and M.W. Shelly, editors, *Multistage Inventory Models and Techniques*, chapter 2, pages 32–46. Stanford University Press, Stanford, CA, 1963.

[22] J. Detar. Semiconductor industry's uncertainty: Sales growth soaring but slowing. 7/6/2010.

[23] A.K. Dixit and R.S. Pindyck. *Investment under Uncertainty. Princeton University Press, Princeton, NJ*, 1994.

[24] K.L. Donohue. Efficient supply contracts for fashion goods with forecast updating and two production modes. *Management Science*, 46(11):1397–1411, 2000.

[25] G.D. Eppen and A.V. Iyer. Improved fashion buying with bayesian updates. *Operations Research*, 45(6):805–819, 1997.

[26] Q. Feng, S.P. Sethi, H. Yan, and H. Zhang. Are base-stock policies optimal in inventory problems with multiple delivery modes? *Operations Research*, 54(4):801–807, 2006.

[27] B. Fleckenstein. Intel's twin woes: Excess capacity, slack demand. 7/19/2004.

[28] I. Friend and M. Blume. The demand for risky assets. *The American Economic Review*, 65(5):900–922, 1975.

[29] Y. Fukuda. Optimal policies for the inventory problem with negotiable lead time. *Management Science*, 10(4):690–708, 1964.

[30] G. Gallego and Ö. Özer. Integrating replenishment decisions with advance demand information. *Management Science*, 47(10):1344–1360, 2001.

[31] N. Golovachkina and J. Bradley. Supplier-manufacturer relationships under forced compliance contracts. *Working Paper, Johnson School of Management, Cornell University*, 2002.

[32] S.C. Graves, H.C. Meal, S. Dasu, and Y. Qui. Two-stage production planning in a dynamic environment. Working paper, Sloan School of Management, MIT, Cambridge, MA, 1986.

[33] S.J. Grossman. Nash equilibrium and the industrial organization of markets with large fixed costs. *Econometrica*, 49:1149–1172, 1981.

[34] C. Haksoz and S. Seshadri. Supply chain operations in the presence of a spot market: a review discussion. *Journal of the Operational Research Society*, 58:1412–1429, 2007.

[35] W.H. Hausman. Sequential decision problems: A model to exploit existing forecast. *Management Science*, 16(2):93–111, 1969.

[36] D.C. Heath and P.L. Jackson. Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. *IIE Transactions*, 26(3):17–30, 1994.

[37] E.F. Hertzler. Intel's capital equipment supply chain. Working paper, Technology Manufacturing Engineering, Intel Corporation, Chandler, AZ, 2009.

[38] D. Heyman and M. Sobel. *Stochastic Models in Operations Research*, volume II. McGraw-Hill, New York, NY, 1984.

[39] G. Hurley, P. Jackson, R. Levi, R.O. Roundy, and D.B. Shmoys. New policies for stochastic inventory control models – theoretical and computational results. submitted. 2007.

[40] Intel Corporation. Intel annual report 2009. Technical report, (Retrieved on July 28, 2010), 2009.

[41] M.E. Johnson. Dual sourcing strategies: Operational hedging and outsourcing to reducing risk in low-cost countries. In C.Y. Lee and H.L. Lee, editors, *Supply Chain Excellence in Emerging Economies*, chapter 5, pages 113–133. Springer Science+Business

Media, New York, NY, 2005.

[42] A. Jones. Prospects for an iron ore cartel. *Resources Policy*, June:103–115, 1986.

[43] B.A. Kalymon. Stochastic prices in a single-item inventory purchasing model. *Operations Research*, 19:1434–1458, 1970.

[44] M. Kanellos. Soaring costs of chipmaking recast industry. *CNET News.com*, `http://news.cnet.com/Semi-survival/2009-10013-981418.html`, 2003.

[45] P.R. Kleindorfer and D. Wu. Integrating long and short-term contracting via business-to-business exchanges for capital intensive industries. *Management Science*, 49(11):1597–1615, 2003.

[46] A.S. Kyle. Informed speculation with imperfect competition. *Review of Economic Studies*, 56, 1989.

[47] R.C. Leachman. The economics of speed. *Fab Engineering and Operations*, 1, 2007.

[48] H. Lee. Rio tinto iron ore: challenges of globalization in the mining industry. *Stanford Global Supply Chain Management Forum - Case GS-56*, 2007.

[49] H. Lee, V. Padmanabhan, and S. Whang. Information distortion in a supply chain: The bullwhip effect. *Management Science*, 43(4):546–558, 1997.

[50] H. Lee and S. Whang. The impact of secondary market on the supply chain. *Management Science*, 48(6):719–731, 2002.

[51] C.L. Li and P. Kouvelis. Flexible and risk-sharing supply contracts under price uncertainty. *Management Science*, 45(10):1378–1398, 1999.

[52] Z. Lian, L. Liu, and S.X. Zhu. Rolling-horizon replenishment: policies and performance analysis. *Naval Research Logistics*, 57:489–502, 2010.

[53] M. Liski and J.P. Montero. Forward trading and collusion in oligopoly. *Journal of Economic Theory*, 131:212–230, 2006.

[54] L. Liu, M. Parlar, and S.X. Zhu. Pricing and leadtime decisions in decentralized supply chains. *Management Science*, 53(5):713–725, 2007.

[55] Q. Liu and G. van Ryzin. Strategic capacity rationing when customers learn. *Manufacturing & Service Operations Management*, 13(1):89–107, 2011.

[56] J. Lovett and S. Anand. The 21st century retailer: Managing customers, merchandise and data. *The Aberdeen Group*, `http://aberdeen.com`, 2007.

[57] H.M. Markowitz. Foundations of portfolio theory. *Journal of Finance*, 46(2):469–477, 1991.

[58] H. Mendelson and T. Tunca. Strategic spot trading in supply chains. *Management Science*, 53(5):742–759, 2007.

[59] S. Minner. Multiple-supplier inventory models in supply chain management: A review. *International Journal of Production Economics*, 81-82:265–279, 2003.

[60] K. Palaka, S. Erlebacher, and D. H. Kropp. Lead time setting, capacity utilization, and pricing decisions under lead time dependent demand. *IIE Transactions*, 30:151–163, 1998.

[61] C. Peng, F. Erhun, E.F. Hertzler, and K.G. Kempf. Capacity planning in the semiconductor industry: Dual-mode procurement with options. Working paper, Department of Management Science and Engineering, Stanford University, Stanford, CA, 2010.

[62] Evan L. Porteus. *Fundations of Stochastic Inventory Theory.* Stanford University Press, 2002.

[63] G. Reinhardt and M. Levesque. A new entrant's decision on virtual versus bricks-and-mortar retailing. *Journal of Electronic Commerce Research*, 5(3):136–152, 2004.

[64] A. Scheller-Wolf and S. Tayur. Managing supply chains in emerging markets. In S. Tayur, R. Ganeshan, and M. Magazine, editors, *Quantitative Models for Supply Chain Management*, chapter 22, pages 703–736. Kluwer Academic Publishers, Norwell, MA, 1998.

[65] A. Scheller-Wolf and S. Tayur. A markovian dual-source production-inventory model with order bands. Working paper, Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, 1998.

[66] N. Secomandi. Optimal commodity trading with a capacitated storage asset. *Management Science*, 56(3):449–467, 2010.

[67] R.W. Seifert, U.W. Thonemann, and W.H. Hausman. Optimal procurement strategies for online spot markets. *European Journal of Operations Research*, 152, 2004.

[68] M. Serapio and N. Trevethan. Interview - singapore's smx plans to launch iron ore futures. *Fox Business*, `http://www.foxbusiness.com/markets/2010/09/28/interview-singapores-smx-plans-launch-iron-ore-futures/`, September 28, 2010.

[69] S.P. Sethi, H. Yan, and H. Zhang. Peeling layers of an onion: Inventory model with multiple delivery modes and forecast updates. *Journal of Optimization Theory and Applications*, 108(2):253–281, 2001.

[70] A. Shapiro. Stochastic programming approach to optimization under uncertainty. *Mathematical Programming*, 112:183–220, 2008.

[71] A.A. Tsay and N. Agrawal. Modeling confict and coordination in multi-channel distribution systems: a review. *Supply Chain Analysis in the E-Business Era*, 2004.

[72] V. Vaidyanathan, D. Metcalf, and D. Martin. Using capacity options to better enable our factory ramps. *Intel Technology Journal*, 9(3):185–191, 2005.

[73] J.A. Van Mieghem. Capacity management, investment, and hedging: Review and recent developments. *Manufacturing & Service Operations Management*, 5(4):269–302, 2003.

[74] A.S. Whittemore and S.C. Saunders. Optimal inventory under stochastic demand with two supply options. *SIAM Journal of Applied Mathematics*, 32:293–305, 1977.

[75] F.A. Wolak. Is price gouging really the problem? *San Diego Union Tribune*, July 27, 2001.

[76] World Bank. Gross domestic product 2009. `http://siteresources.worldbank.org/DATASTATISTICS/Resources/GDP.pdf`, (Retrieved on July 28, 2010), 2009.

[77] D.J. Wu, P. Kleindorfer, and J.E. Zhang. Optimal bidding and contracting strategies for capital-intensive goods. *European Journal of Operations Research*, 137:657–676, 2002.

[78] S.D. Wu, M. Erkoc, and S. Karabuk. Managing capacity in the high-tech industry: A review of literature. *The Engineering Economist*, 50:125–158, 2005.

[79] S.D. Wu, K.G. Kempf, M.O. Atan, B. Aytac, S.A. Shirodkar, and A. Mishra. Improving new-product forecasting at intel corporation. *Interfaces (to appear)*, Wagner Prize issue, 2010.

[80] H. Yan, K. Liu, and A. Hsu. Optimal ordering in a dual-supplier system with demand forecast updates. *Production and Operations Management*, 12(1):30–45, 2003.

[81] Ö. Yazlali and F. Erhun. Relating the multiple supply problem to quantity flexibility contracts. *Operations Research Letters*, 35(6):767–772, 2007.

[82] Ö. Yazlali and F. Erhun. Dual-supply inventory problem with capacity limits on order sizes and unrestricted ordering costs. *IIE Transactions*, 41(8):716–729, 2009.

[83] Ö. Yazlali and F. Erhun. Managing demand uncertainty with dual supply contracts on capacity and inventory: A heuristic approach. Working paper, Department of Management Science and Engineering, Stanford University, Stanford, CA, 2010.

[84] I. Yildirim, B. Tan, and F. Karaesmen. A multiperiod stochastic production planning and sourcing problem with service level constraints. *OR Spectrum*, 27:471–489, 2005.